THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XIII

1942

THE ANNALS OF MATHEMATICAL STATISTICS

The Annals of Mathematical Statistics is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the Annals of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the Annals of Mathematical Statistics should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the Annals is \$5.00 per year. Single copies \$1.50. Back numbers are available at \$5.00 per volume, or \$1.50 per single issue.

Composed and Printed at the
WAVERLY PRESS, Inc.
Baltimore, Md., U. S. A.

CONTENTS OF VOLUME XIII

Articles

NOTES

MISCELLANEOUS

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIII, No. 1 — March, 1942

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Ave., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT

By R. L. Anderson

North Carolina State College

1. Introduction. The problem of serial correlation was brought to the attention of statisticians by Yule in 1921 [9] Both Yule and Bartlett [2] have shown that the ordinary tests of significance are invalidated if successive observations are not independent of one another. The serial correlation coefficient has been introduced as a measure of the relationship between successive values of a variable ordered in time or space. Interest in the serial correlation problem was stimulated further by the new concepts of time series analysis discussed by Wold [8].

We shall define the serial correlation coefficient for lag L and N observations to be

$$ {}_LR_N = \frac{{}_LC_N}{V_N} = \frac{X_1X_{L+1} + X_2X_{L+2} + \cdots + X_NX_L - (\Sigma X_i)^2/N}{\Sigma X_i^2 - (\Sigma X_i)^2/N}, $$

where C and V are the covariance and variance respectively and the X's are considered to be independently normally distributed about the same mean with unit variance.[1] If the population variance were known a priori, the variates could be transformed so that they would have unit variance; under such an unusual circumstance, the only distribution required would be that of the serial covariance. Tintner has given a test of significance for the serial covariance [6] and for the correlation coefficient [7] by using a method of selected items. The author has presented the distribution of the serial covariance and of the serial correlation coefficient not corrected for the mean in a recent doctoral thesis [1]. The distributions of ${}_LR_N$ not corrected for the mean will be mentioned in the sections which follow.

2. Small sample distributions for lag 1. W. G. Cochran has suggested that we use a result given in his article on quadratic forms to derive the distributions of the serial correlation coefficient for small samples [3]. If $X_1, X_2, \cdots, X_N$ are independently normally distributed with variance 1 and mean 0, then

> "Every quadratic form $\Sigma a_{ij}X_iX_j$ is distributed like $\sum_{k=1}^{r} \lambda_k u_k$, where r is the rank of the matrix, A, of the quadratic form, the u's are independently distributed as χ^2, each with 1 d.f., and the λ's are the non-zero latent roots of the characteristic equation of A" [3, p. 179].

If each λ_i appears k_i times as a latent root, u_i will be distributed as χ^2 with k_i degrees of freedom.

[1] This circular definition of the serial correlation coefficient was suggested by H. Hotelling.

If we set $L = 1$ in the above definition of the serial covariance, we note that the characteristic equation of ${}_1C_N$ is

$$ {}_1F_N = \begin{vmatrix} a_1 & a_2 & a_3 & \cdots & a_N \\ a_N & a_1 & a_2 & \cdots & a_{N-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{vmatrix} = 0, $$

where $a_1 = -(\lambda + 1/N)$, $a_2 = a_N = (N-2)/2N$, and all other a's $= -1/N$. The determinant can be evaluated by the method of circulants. We find that ${}_1F_N(\lambda) = \prod_{K=1}^{N}\left\{\sum_{i=1}^{N} a_i\omega_k^{i-1}\right\}$, where ω_k is the kth root of unity. Hence,

$$ {}_1F_N = \prod_{K=1}^{N}\left\{-\left(\lambda_k + \frac{1}{N}\right) + \frac{N-2}{2N}(\omega_k + \omega_k^{-1}) - \frac{1}{N}\sum_{i=3}^{N-1}\omega_k^{i-1}\right\}. $$

Since

$$ \sum_{i=3}^{N-1}\omega_k^{i-1} = \begin{cases} -(\omega_k + 1 + \omega_k^{-1}), & \text{for } k \neq N \\ (N-3), & \text{for } k = N \end{cases} $$

$$ {}_1F_N = \prod_{K=1}^{N-1}\{-\lambda_k + (\omega_k + \omega_k^{-1})/2\} = \prod_{K=1}^{N-1}\left\{-\lambda_k + \cos\frac{2\pi k}{N}\right\} = 0. $$

Hence $\lambda_k = \cos\frac{2\pi k}{N}$, $(k = 1, 2, \cdots, N-1)$, and

$$ {}_1C_N = \begin{cases} \sum_{K=1}^{\frac{1}{2}(N-1)} \lambda_k u_k, & \text{for } N \text{ odd}, \\ \sum_{K=1}^{\frac{1}{2}(N-2)} \lambda_k u_k - u, & \text{for } N \text{ even}, \end{cases} $$

where u_k is distributed as χ^2 with 2 d.f. and u with 1 d.f. At the same time, we note that $V_N = \Sigma(X_i - \bar{X})^2$ is distributed as χ^2 with $N-1$ d.f.

The general procedure in deriving the distribution of ${}_1R_N$ is as follows: We determine the joint density function of the u's which form the distributions of ${}_1C_N(= {}_1R_N \cdot V_N)$ and V_N. The u's are integrated out, leaving the joint density function of ${}_1R_N$ and V_N. The distribution of ${}_1R_N$ is obtained by integrating with respect to V_N from 0 to ∞. As examples, derivations of the distributions of ${}_1R_6$ and ${}_1R_7$ have been included. In order to simplify the results, the first subscripts have been dropped from ${}_1R_N$.

Distribution of R_6. $R_6V_6 = \lambda_1u_1 + \lambda_2u_2 - u$ and $V_6 = u_1 + u_2 + u$, where u_1 and u_2 are distributed as χ^2 with 2 d.f. and u with 1 d.f. and $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = -\frac{1}{2}$. Hence the density function of the u's is

$$ D(u_1, u_2, u) = (4\sqrt{2\pi})^{-1}u^{-\frac{1}{2}}e^{-\frac{1}{2}V_6}. $$

Since $u_1 = [V_6(R_6 - \lambda_2) + u(1 + \lambda_2)]/(\lambda_1 - \lambda_2)$ and

$$u_2 = [V_6(\lambda_1 - R_6) - u(1 + \lambda_1)]/(\lambda_1 - \lambda_2),$$

u must vary between 0 and $V_6(\lambda_1 - R_6)/(1 + \lambda_1)$ for $\lambda_2 \leq R_6 \leq \lambda_1$ and between $V_6(\lambda_2 - R_6)/(1 + \lambda_2)$ and $V_6(\lambda_1 - R_6)/(1 + \lambda_1)$ for $-1 \leq R_6 \leq \lambda_2$. After integrating with respect to u between these limits and then with respect to V_6 from 0 to ∞, we obtained the following density function for R_6:

$$D(R_6) = \frac{3}{2}\begin{cases} \dfrac{\sqrt{(\lambda_1 - R_6)}}{\sqrt{(1 + \lambda_1)}\,(\lambda_1 - \lambda_2)}, & \text{for } \lambda_2 \leq R_6 \leq \lambda_1 \\[2ex] \dfrac{\sqrt{(\lambda_1 - R_6)}}{\sqrt{(1 + \lambda_1)}\,(\lambda_1 - \lambda_2)} + \dfrac{\sqrt{(\lambda_2 - R_6)}}{\sqrt{(1 + \lambda_2)}\,(\lambda_2 - \lambda_1)}, & \text{for } -1 \leq R_6 \leq \lambda_2. \end{cases}$$

The cumulative probability function has the same general form:

$$P(R_6 > R') = \begin{cases} \dfrac{(\lambda_1 - R')^{\frac{3}{2}}}{\sqrt{(1 + \lambda_1)}\,(\lambda_1 - \lambda_2)} + \dfrac{(\lambda_2 - R')^{\frac{3}{2}}}{\sqrt{(1 + \lambda_2)}\,(\lambda_2 - \lambda_1)} & \text{for } -1 \leq R' \leq \lambda_2 \\[2ex] \dfrac{(\lambda_1 - R')^{\frac{3}{2}}}{\sqrt{(1 + \lambda_1)}\,(\lambda_1 - \lambda_2)} & \text{for } \lambda_2 \leq R' \leq \lambda_1 \end{cases}$$

Distribution of R_7. $R_7V_7 = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3$ and $V_7 = u_1 + u_2 + u_3$, where each u is distributed as χ^2 with 2 d.f. Hence,

$$u_1 = \frac{V_7(R_7 - \lambda_2) + u_3(\lambda_2 - \lambda_3)}{(\lambda_1 - \lambda_2)} \quad \text{and} \quad u_2 = \frac{V_7(\lambda_1 - R_7) - u_3(\lambda_1 - \lambda_3)}{(\lambda_1 - \lambda_2)}.$$

For $\lambda_2 \leq R_7 \leq \lambda_1$, $0 \leq u_3 \leq V_7(\lambda_1 - R_7)/(\lambda_1 - \lambda_3)$; for $\lambda_3 \leq R_7 \leq \lambda_2$, $V_7(\lambda_2 - R_7)/(\lambda_2 - \lambda_3) \leq u_3 \leq V_7(\lambda_1 - R_7)/(\lambda_1 - \lambda_3)$. Using these limits, we derived the following density function for R_7:

$$D(R_7) = 2 \cdot \begin{cases} \dfrac{(\lambda_1 - R_7)}{(\lambda_1 - \lambda_2)\,(\lambda_1 - \lambda_3)} + \dfrac{(\lambda_2 - R_7)}{(\lambda_2 - \lambda_1)\,(\lambda_2 - \lambda_3)} & \text{for } \lambda_3 \leq R_7 \leq \lambda_2 \\[2ex] \dfrac{(\lambda_1 - R_7)}{(\lambda_1 - \lambda_2)\,(\lambda_1 - \lambda_3)} & \text{for } \lambda_2 \leq R_7 \leq \lambda_1. \end{cases}$$

The cumulative probability function is similar, except that the coefficient 2 cancels and the exponent of each numerator is raised by one.

General formulas for N odd. It appears that the density function for R_N and V_N for N odd is

$$D(R_N, V_N) = KV_N^{\frac{1}{2}(N-3)}e^{-\frac{1}{2}V_N}\sum_{i=1}^{m}(\lambda_i - R_N)^{\frac{1}{2}(N-5)}/\alpha_i \qquad \text{for } \lambda_{m+1} \leq R_N \leq \lambda_m,^{2}$$

where $\alpha_i = \prod_{j=1}^{\frac{1}{2}(N-1)}{}' (\lambda_i - \lambda_j)$ for $j \neq i$ and $1/K = 2^{\frac{1}{2}(N-1)}\Gamma[\frac{1}{2}(N-3)]$. This

[2] Note that we are omitting the lag subscript from ${}_1R_N$.

formula holds for $N = 5$ and 7; we will show that it holds for $N + 2$, assuming it true for N. If we set $k = \frac{1}{2}(N + 1)$, $R_{N+2}V_{N+2} = R_N V_N + \lambda_k u_k$ and $V_{N+2} = V_N + u_k$; hence,

$$R_N = \frac{(R_{N+2}V_{N+2} - \lambda_k u_k)}{V_{N+2} - u_k} \quad \text{and} \quad V_N = V_{N+2} - u_k .$$

If we make the substitution $u_k = u_k' V_{N+2}$, the density function for u_k', V_{N+2}, and R_{N+2} is

$$\tfrac{1}{2}KV_{N+2}^{\frac{1}{2}(N-1)} e^{-\frac{1}{2}V_{N+2}} \sum_{i=1}^{m} [(\lambda_i - R_{N+2}) - u_k'(\lambda_i - \lambda_k)]^{\frac{1}{2}(N-3)}/\alpha_i .$$

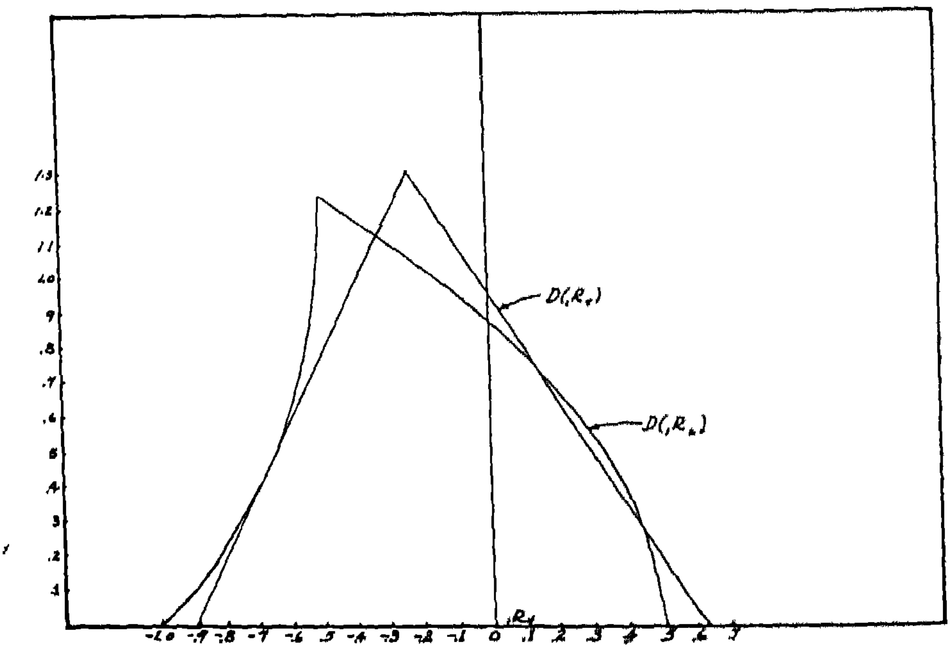


FIG. 1

In order to obtain the distribution of V_{N+2} and R_{N+2}, we must integrate out u_k'. The limits of integration differ for different values of m. We note that

$$u_k' = (R_N - R_{N+2})/(R_N - \lambda_k),$$

except that $u_k' \equiv 0$ when $\lambda_k < R_N \leq \lambda_{m+1}$, since $\lambda_{m+1} \leq R_{N+2} \leq \lambda_m$ and u_k' can not be negative. For $R_{N+2} > \lambda_k$, $u_k' < 1$; hence, if R_N is replaced by a larger (smaller) quantity, u_k' will be larger (smaller).

For $m = 1(\lambda_2 \leq R_{N+2} \leq \lambda_1)$, we need to consider only that region for which $\lambda_2 \leq R_N \leq \lambda_1$. In this region, $0 \leq u_k' \leq (\lambda_1 - R_{N+2})/(\lambda_1 - \lambda_k)$ and the density function of R_{N+2} and V_{N+2} is

$$\phi(V_{N+2})(\lambda_1 - R_{N+2})^{\frac{1}{2}(N-3)}/\alpha_1',$$

where $\phi(V_{N+2}) = V_{N+2}^{\frac{1}{2}(N-1)} e^{-\frac{1}{2}V_{N+2}}/2^{\frac{1}{2}(N+1)} \cdot \Gamma[\frac{1}{2}(N-1)]$ and $\alpha_1' = \prod_{j=2}^{K} (\lambda_1 - \lambda_j)$.

For $m = 2(\lambda_3 \le R_{N+2} \le \lambda_2)$, we must consider two regions in the R_N plane. When $\lambda_2 \le R_N \le \lambda_1$,

$$\frac{\lambda_2 - R_{N+2}}{\lambda_2 - \lambda_k} \le u_k' \le \frac{\lambda_1 - R_{N+2}}{\lambda_1 - \lambda_k},$$

and when $\lambda_3 \le R_N \le \lambda_2$, $0 \le u_k' \le (\lambda_2 - R_{N+2})/(\lambda_2 - \lambda_k)$. If we combine the density functions for these two regions, we find that

$$D(R_{N+2}, V_{N+2}) = \phi(V_{N+2}) \sum_{i=1}^{2} (\lambda_i - R_{N+2})^{\frac{1}{2}(N-3)}/\alpha_i' \qquad \text{for } \lambda_3 \le R_{N+2} \le \lambda_2 .$$

Similar results can be obtained for the other regions.

Finally we conclude that for N odd,

$$D({}_1R_N) = \tfrac{1}{2}(N-3) \sum_{i=1}^{m} (\lambda_i - {}_1R_N)^{\frac{1}{2}(N-5)}/\alpha_i \qquad \text{for } \lambda_{m+1} \le {}_1R_N \le \lambda_m$$

and

$$P({}_1R_N > R') = \sum_{i=1}^{m} (\lambda_i - R')^{\frac{1}{2}(N-3)}/\alpha_i \qquad \text{for } \lambda_{m+1} \le R' \le \lambda_m ,$$

where $\alpha_i = \prod_{j=1}^{\frac{1}{2}(N-1)}{}' (\lambda_i - \lambda_j)$, $i \ne j$. The general density function for N odd and ${}_1R_N$ not corrected for the sample mean is [1]

$$D({}_1R_N) = \tfrac{1}{2}(N-2) \sum_{i=m}^{\frac{1}{2}(N-1)} ({}_1R_N - \lambda_i)^{\frac{1}{2}(N-4)}/\alpha_i \quad \text{for} \quad \lambda_m \le {}_1R_N \le \lambda_{m-1},$$

where $\alpha_i = \prod_{j=1}^{\frac{1}{2}(N-1)}{}' (\lambda_j - \lambda_i)\sqrt{(1 - \lambda_i)}$, $i \ne j$.

General formulas for N even. Using the same method as above, we can show that the same formulas hold for N even and ${}_1R_N$ corrected for the mean except that in this case $\alpha_i = \prod_{j=1}^{\frac{1}{2}(N-2)}{}' (\lambda_i - \lambda_j)\sqrt{(\lambda_i + 1)}$, $j \ne i$. No general formulas were derived for N even and ${}_1R_N$ not corrected for the mean.

3. Large sample distributions for lag 1. The simultaneous density function of C and V, where we will drop the subscripts for convenience, is

$$D(C, V) = (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(s, t) e^{-isC - itV} \, ds \, dt.$$

$$\phi(s, t) = K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta} \, dX_1 \, dX_2 \cdots dX_N ,$$

where $\theta = \{\Sigma X_i^2 - 2t[\Sigma(X_i - \bar{X})^2] - 2s[X_1X_2 + \cdots + X_NX_1 - (\Sigma X_i)^2/N]\}$ and s and t are pure imaginaries.

$\phi(s, t) = \Delta^{-\frac{1}{2}}$, where Δ is the determinant of the quadratic form θ. This determinant was evaluated by the method of circulants; we found that $\Delta = \prod_{k=1}^{N-1} \{1 - 2(t + s\lambda_k)\}$, where $\lambda_k = \cos 2\pi k/N$.

Set $K = \log \phi(s, t) = \Sigma \kappa_{ij} \frac{s^i t^j}{i!j!}$. If K is expanded in series, we find that $\kappa_{ij} = m!2^m \sum_{k=1}^{N-1} \lambda_k^i$, where $m = (i + j - 1)$. For $N > i$, we might evaluate these summations: $\Sigma\lambda_k = -1$, $\Sigma\lambda_k^2 = \frac{1}{2}(N - 2)$, $\Sigma\lambda_k^3 = -1$, $\Sigma\lambda_k^4 = \frac{1}{8}(3N - 8)$ and $\Sigma\lambda_k^5 = -1$. Hence $\kappa_{10} = E(C) = -1$, $\kappa_{01} = E(V) = (N - 1)$, $\kappa_{20} = \sigma_c^2 = 2(N - 2)$, $\kappa_{02} = \sigma_v^2 = 2(N - 1)$, $\kappa_{11} = \rho\sigma_c\sigma_v = -2$, $\kappa_{30} = -8$, $\kappa_{03} = 8(N - 1)$, $\kappa_{21} = 4(N - 2)$, $\kappa_{12} = -8$, etc.

If we let $C' = C + 1$ and $V' = V - (N - 1)$, all of these semi-invariants will remain unchanged except that $\kappa_{10} = \kappa_{01} = 0$. Since $R = C/V$,

$$\left(R + \frac{1}{N-1}\right) = \frac{C'(N-1) + V'}{[V' + (N-1)](N-1)}$$

$$= \frac{C'(N-1) + V'}{(N-1)^2} \left\{\sum_{p=0}^{\infty} (-1)^p \left(\frac{V'}{N-1}\right)^p\right\}$$

If we neglect terms of order less than $1/N$, $E(R) = -1/(N - 1)$, $E(R - \bar{R})^2 = \frac{(N-2)}{(N-1)^2}$, and $E(R - \bar{R})^k = 0$ for $k > 2$. For $N < 75$, a more exact approximation may be desired.

If the above approximation is used, ${}_1R_N$ is normally distributed with mean $-1/(N - 1)$ and variance $(N - 2)/(N - 1)^2$. The single-tail significance points can be found by substituting in the formulas

$${}_1R_N(.05) = \frac{-1 \pm 1.645\sqrt{(N-2)}}{N-1} \quad \text{or} \quad {}_1R_N(.01) = \frac{-1 \pm 2.326\sqrt{(N-2)}}{N-1}$$

Refer to Fig. 2 for a comparison of the exact distribution and the normal approximation for $N = 15$. I have included the graphs of the exact distributions for $N = 6$ and 7 in Fig. 1. We might note a few comparisons between the approximate significance points and the exact ones:

N	Positive tail				Negative tail			
	5%		1%		5%		1%	
	Exact	Approx.	Exact	Approx.	Exact	Approx.	Exact	Approx.
45	0.218	0.223	0.314	0.324	−0.262	−0.268	−0.350	−0.369
75	0.173	0.176	0.250	0.255	−0.199	−0.203	−0.276	−0.282

For ${}_1R_N$ not corrected for the mean, it was found that $y = \sqrt{\dfrac{N\,{}_1R_N^2}{1 + 2\,{}_1R_N^2}}$ was asymptotically normally distributed with mean 0 and variance 1 [1].

4. **Significance points of ${}_1R_N$.** An example of the methods used in tabulating these significance points has been presented in the author's doctoral thesis [1]. The significance points for the values of N enclosed in parentheses have been obtained by graphical interpolation. Note that N is the number of observations (see Table I).

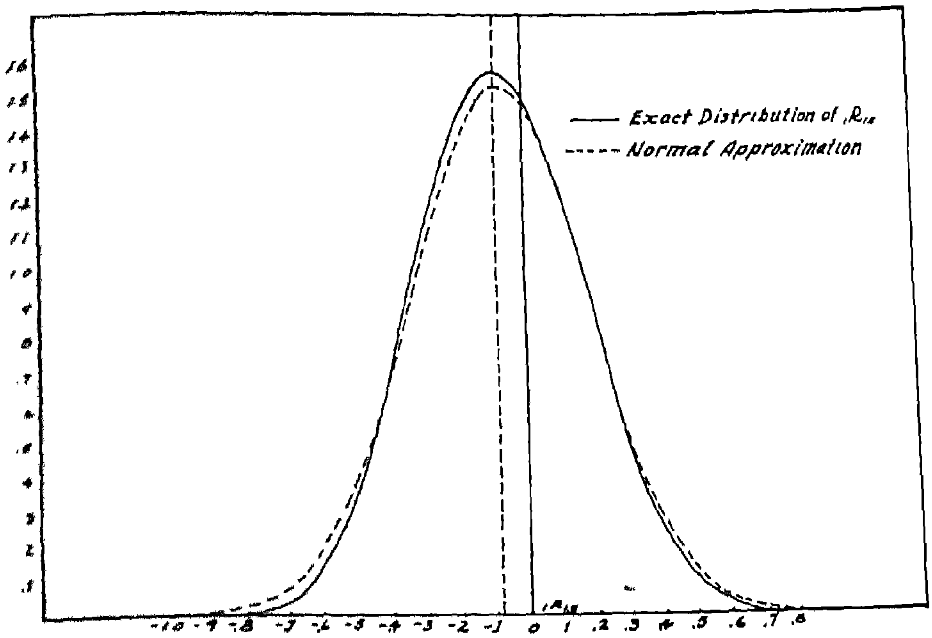


FIG. 2

5. **Distributions for general lag, L.** (a) *Introduction.* For a general lag, L, the constants in the characteristic equation for the covariance ${}_LC_N$ are $a_1 = -(\lambda + 1/N)$, $a_{L+1} = a_{N-L+1} = (N-2)/2N$ and all other a's $= -1/N$. Hence the characteristic equation is

$$ {}_LF_N = \prod_{k=1}^{N-1} [\lambda_k - \cos(2\pi Lk/N)] = 0. $$

Certain important generalizations concerning ${}_LF_N$ may be set down:

1. When L is not a factor of N or has no common factor with N, ${}_LF_N = {}_1F_N$.
2. When L and N have a common factor, α, ${}_LF_N = ({}_1F_{N/\alpha})^\alpha(\lambda - 1)^{\alpha-1}$.

2a. If $\alpha = L$, ${}_LF_N = ({}_1F_p)^L(\lambda - 1)^{L-1}$, where $p = N/L$.

The proof of the first statement was suggested by Cochran. Since $\cos(\alpha + 2a\pi) = \cos\alpha$, where a is any integer we must prove that the series of numbers

$$L, 2L, \cdots, (N - 1)L,$$

when reduced modulus N can be arranged to form the series

$$1, 2, \cdots, (N - 1).$$

This proof can be found in most books on the theory of numbers; e.g. [4]. Hence we conclude that each term of the sequence $\{\cos (2\pi Lk/N)\}$ reduces uniquely

TABLE I

N	Positive tail		Negative tail	
	5%	1%	5%	1%
5	0.253	0.297	−0.753	−0.798
6	0.345	0.447	0.708	0.863
7	0.370	0.510	0.674	0.799
8	0 371	0.531	0.625	0.764
9	0.366	0.533	0.593	0.737
10	0.360	0.525	0.564	0.705
11	0.353	0.515	0.539	0.679
12	0.348	0.505	0.516	0.655
13	0.341	0.495	0.497	0.634
14	0.335	0.485	0.479	0.615
15	0.328	0.475	0.462	0.597
20	0.299	0.432	0.399	0.524
25	0.276	0.398	0.356	0.473
30	0.257	0.370	0.325	0.433
(35)	0.242	0.347	0.300	0.401
(40)	0.229	0.329	0.279	0.376
45	0.218	0.314	0.262	0.356
(50)	0.208	0.301	0.248	0.339
(55)	0.199	0.289	0.236	0.324
(60)	0.191	0.278	0.225	0.310
(65)	0.184	0.268	0.216	0.298
(70)	0.178	0.259	0.207	0.287
75	0 173	0.250	−0.199	−0.276

to one of the sequence $\{\cos (2\pi k/N)\}$ for $k = 1, 2, \cdots, (N - 1)$, when L/N is a prime fraction.

If L and N have a common factor, α, $L = q\alpha$ and $N = p\alpha$, where p and q are integers prime to one another. Hence,

$$\begin{aligned} {}_LF_N &= \prod_{k=1}^{p\alpha-1} \left\{\lambda_k - \cos \frac{2\pi qk}{p}\right\} = \prod_{k=1}^{p-1} \left(\lambda_k - \cos \frac{2\pi k}{p}\right)^{\alpha} (\lambda - \cos 2\pi)^{\alpha-1} \\ &= ({}_1F_p)^{\alpha}(\lambda - 1)^{\alpha-1} = 0. \end{aligned}$$

If $\alpha = L$, ${}_LF_N = ({}_1F_p)^L(\lambda - 1)^{L-1}$, where $p = N/L$.

When these results are applied to the large sample distribution of ${}_LR_N$, we find that it is independent of L. For the more important case in which $p = N/L$, the semi-invariants κ_{ij} for C and V are exactly the same for all L with a given N. We see that

$$K_L = -\tfrac{1}{2}L \sum_{k=1}^{p-1} \log \{1 - 2(t + s\lambda_k)\} - \tfrac{1}{2}(L - 1) \log \{1 - 2(t + s)\},$$

where $\lambda_k = \cos (2\pi k/p)$. Hence, $\kappa_{ij} = m!2^m \left\{ \frac{N}{p} \left(\sum_{k=1}^{p-1} \lambda_k^i + 1 \right) - 1 \right\}$. But $\sum_{k=1}^{p-1} \lambda_k^i + 1$ is always 0 or a multiple of p when $p > i$; therefore, the p's cancel and κ_{ij} is the same for all p or for all L, since $L = N/p$. When $p \leq i$, the κ_{ij}'s will not be equal for all p. For example $\kappa_{20} = 2(N - 1)$ for $p = 2$ and $\kappa_{30} = 2(N - 4)$ for $p = 3$.

(b) *Distributions of ${}_LR_N$ when $N/L = p$.* These results indicate that the distributions of the serial correlation coefficients for which the number of observations is divisible by the lag, so that $N/L = p$, would include the distributions of all the serial correlation coefficients regardless of the values of N and L. We will designate any lag L as the primary lag for a given N if $N/L = p$, an integer. For example, ${}_2R_6$ and ${}_4R_8$ have the same density function, but we will derive only the density function for lag 2, which we will call the primary lag. The case of $p = 1$ is trivial, since it involves correlating a series with itself. To date, we have derived the exact density functions for $p = 2$ and $p = 3$ and the required integrals for $p = 4$. The significance points have been tabulated in Table II. For simplicity of notation, we will set ${}_LR_N = {}_LR_p$ and $V_N = V$.

Case $p = 2(N = 2L)$. ${}_LR_2V = -u_1 + u_2$ and $V = u_1 + u_2$, where u_1 is distributed as χ^2 with L d.f. and u_2 as χ^2 with $L - 1$ d.f. Hence,

$$D_L(u_1, u_2) = K(u_1)^{\frac{1}{2}(L-2)}(u_2)^{\frac{1}{2}(L-3)},$$

where $1/K = 2^{L-\frac{1}{2}}\Gamma(\tfrac{1}{2}L)\Gamma[\tfrac{1}{2}(L - 1)]e^{V/2}$. After substituting $u_1 = V(1 - {}_LR_2)/2$ and $u_2 = V(1 + {}_LR_2)/2$ and integrating with respect to V from 0 to ∞, we have

$$D({}_LR_2) = \frac{(1 - {}_LR_2)^{\frac{1}{2}(L-2)}(1 + {}_LR_2)^{\frac{1}{2}(L-3)}}{2^{L-\frac{3}{2}}\beta[\tfrac{1}{2}L, \tfrac{1}{2}(L - 1)]}.$$

If we set $(1 - {}_LR_2) = 2y$, then the cumulative probability function is

$$P({}_LR_2 > R') = \frac{1}{\beta[\tfrac{1}{2}L, \tfrac{1}{2}(L - 1)]} \int_{y=0}^{\frac{1}{2}(1-R')} y^{\frac{1}{2}(L-2)}(1 - y)^{\frac{1}{2}(L-3)}\, dy.$$

Pearson has tabulated the values of these incomplete Beta functions [5]. In his notation, $P = I_x[\tfrac{1}{2}L, \tfrac{1}{2}(L - 1)]$, where $x = \tfrac{1}{2}(1 - R')$. For ${}_LR_2$ not corrected for the mean, $P = I_x(\tfrac{1}{2}L, \tfrac{1}{2}L)$ [1].

Case $p = 3(N = 3L)$. ${}_LR_3V = -\tfrac{1}{2}u_1 + u$ and $V = u_1 + u$, where u_1 is distributed as χ^2 with $2L$ d.f. and u with $L - 1$ d.f. Therefore, $D_L(u_1, u) =$

$Ku_1^{L-1}u^{\frac{1}{2}(L-3)}$, where $1/K = 2^{\frac{1}{2}(3L-1)}\Gamma(L)\Gamma[\frac{1}{2}(L-1)]e^{V/2}$. After substituting $u_1 = 2V(1 - {}_LR_3)/3$ and $u = V(1 + 2{}_LR_3)/3$ and integrating with respect to V from 0 to ∞, we find that

$$D({}_LR_3) = \frac{2^L(1 - {}_LR_3)^{L-1}(1 + 2\,{}_LR_3)^{\frac{1}{2}(L-3)}}{3^{\frac{1}{2}(L-1)}\beta[L, \frac{1}{2}(L-1)]}, \qquad {}_LR_3 \geq -\tfrac{1}{2}.$$

If we set $x = 2(1 - R')/3$, $P({}_LR_3 > R') = I_x[L, \frac{1}{2}(L-1)]$. For ${}_LR_3$ not corrected for the mean, $P = I_x[L, \frac{1}{2}L]$.

Case $p = 4(N = 4L)$. ${}_LR_4V = -u_2 + u_4$ and $V = u_2 + u_4 + u$, where u_2 is distributed as χ^2 with L d.f., u_4 with $L - 1$ d.f. and u with $2L$ d.f. The density function of the u's is $D_L(u_2, u_4, u) = Ku_2^{\frac{1}{2}(L-2)}u_4^{\frac{1}{2}(L-3)}u^{L-1}e^{-V/2}$, where $1/K = 2^{\frac{1}{2}(4L-1)}\Gamma(\frac{1}{2}L)\Gamma[\frac{1}{2}(L-1)]\Gamma(L)$. Since $u_4 = [V(1 + {}_LR_4) - u]/2$ and $u_2 = [V(1 - {}_LR_4) - u]/2$, $0 \leq u \leq V(1 - {}_LR_4)$ for ${}_LR_4 \geq 0$ and $0 \leq u \leq V(1 + {}_LR_4)$ for ${}_LR_4 \leq 0$. For ${}_LR_4 \geq 0$,

$$D({}_LR_4) = \frac{KVe^{-\frac{1}{2}V}}{2^{\frac{1}{2}(2L-3)}} \int_{u=0}^{V(1-{}_LR_4)} [V(1 + {}_LR_4) - u]^{\frac{1}{2}(L-3)}[V(1 - {}_LR_4) - u]^{\frac{1}{2}(L-2)}u^{L-1}\,du.$$

For ${}_LR_4 \leq 0$, $D({}_LR_4)$ is the same except that the upper limit for the integral is $V(1 + {}_LR_4)$. If we make the substitution $y = u/(\text{upper limit})$ in each case and then integrate with respect to V from 0 to ∞, we have these density functions:

$$D({}_LR_4) = k\cdot\begin{cases}(1 + {}_LR_4)^{\frac{1}{2}(3L-3)}\displaystyle\int_{y=0}^{1} y^{L-1}(1-y)^{\frac{1}{2}(L-3)}[(1 - {}_LR_4) - y(1 + {}_LR_4)]^{\frac{1}{2}(L-2)}\,dy, \\ \hfill \text{for } {}_LR_4 \leq 0, \\ (1 - {}_LR_4)^{\frac{1}{2}(3L-2)}\displaystyle\int_{y=0}^{1} y^{L-1}(1-y)^{\frac{1}{2}(L-2)}[(1 + {}_LR_4) - y(1 - {}_LR_4)]^{\frac{1}{2}(L-3)}\,dy, \\ \hfill \text{for } {}_LR_4 \geq 0,\end{cases}$$

where $k = \Gamma[\frac{1}{2}(4L-1)]/2^{\frac{1}{2}(2L-3)}\cdot\Gamma(L)\cdot\Gamma(\frac{1}{2}L)\cdot\Gamma[\frac{1}{2}(L-1)]$.

The probability integrals must be evaluated for each L. The cumulative probability functions for $L = 2$ and 3 are:

$$P({}_LR_4 > R') = 1 - \frac{\sqrt{2}}{2}\cdot\begin{cases}(1 + R')^{5/2} - R'^{3/2}(5 + R')/\sqrt{2}, & \text{for } R' \geq 0, \\ (1 + R')^{5/2}, & \text{for } R' \leq 0,\end{cases}$$

$$P({}_LR_4 > R') = \frac{\sqrt{2}}{4}\begin{cases}(1 - R')^{9/2}, & \text{for } R' \geq 0, \\ (1 - R')^{9/2} - (-R'/2)^{5/2}(22R'^2 + 36R' + 126), & \text{for } R' \leq 0.\end{cases}$$

Since the density functions are much simpler for $R' > 0$ when L is odd and for $R' < 0$ when L is even, we have derived only these significance points for $L > 3$ and interpolated for the intermediate points. It was noted that the significance points approach those given in Table I for the first lag. For these comparisons, see Table III below. Note that for $L \geq 7$ the 5% points are almost identical and the 1% points are nearly accurate to two decimal places.

TABLE II

Significance points of $_LR_N$ for $p = 2$ and 3[3]

L[3]	p=2 ($N=2L$)				p=3 ($N=3L$)			
	Positive tail		Negative tail		Positive tail		Negative tail	
	5%	1%	5%	1%	5%	1%	5%	1%
2	0.805	0.960	−0.99	−1.00	0.488	0.762	−0.496	−0.50
3	0.729	0.907	0.928	0.994	0.447	0.677	0.474	0.490
4	0.664	0.852	0.848	0.950	0.406	0.610	0.439	0.480
5	0.612	0.802	0.773	0.902	0.373	0.559	0.406	0.461
6	0.571	0.759	0.712	0.856	0.346	0.518	0.377	0.440
7	0.536	0.721	0.662	0.812	0.324	0.485	0.354	0.420
8	0.507	0.688	0.620	0.774	0.306	0.457	0.334	0.402
9	0.483	0.659	0.585	0.739	0.291	0.433	0.316	0.387
10	0.462	0.634	0.554	0.708	0.278	0.413	0.301	0.373
12	0.428	0.590	0.505	0.656	0.256	0.380	0.276	0.347
14	0.399	0.554	0.467	0.612	0.239	0.353	0.256	0.326
16	0.376	0.523	0.436	0.577	0.225	0.332	0.240	0.308
18	0.357	0.498	0.410	0.546	0.213	0.314	0.227	0.293
20	0.340	0.476	0.389	0.520	0.202	0.298	0.215	0.280
25	0.308	0.432	0.347	0.469	0.182	0.268	0.193	0.254
30	0.282	0.398	0.317	0.431	0.167	0.245	0.176	0.234
40	0.247	0.348	0.273	0.374	0.146	0.212	0.153	0.205
50	0.222	0.314	−0.243	−0.335	0.131	0.191	−0.136	−0.184

TABLE III[4]

Significance points for $p = 4$

L	N	Positive tail				Negative tail			
		5%		1%		5%		1%	
		Exact	Table 1	Exact	Table 1	Exact	Table 1	Exact	Table 1
2	8	0.373	0.371	0.618	0.531	−0.653	−0.625	−0.818	−0.764
3	12	0.353	0.348	0.547	0.505	0.528	0.516	0.692	0.655
4	16	0.325*	0.322	0.490*	0.466	0.451	0.447	0.604	0.580
5	20	0.301	0.299	0.451	0.432	0.402*	0.409	0.543*	0.524
6	24	0.281*	0.280	0.419*	0.404	0.365	0.363	0.497	0.482
7	28	0.264	0.264	0.392	0.380	−0.338*	−0.337	−0.460*	−0.448

[3] L is the lag and $p = N/L$.
[4] * indicates interpolated values.

Case $p > 4$. We have not set up any of the density functions for $p > 4$; however, it appears that the significance points given for lag 1 would be accurate enough for the higher lags. The exact significance points for lag 2 have been derived for $p = 5$ and 7. The reader may note the close approximation given by the significance points for lag 1 when $p = 7$. We hope to check the lag 1 approximation for other lags in the near future.

TABLE IV

Some significance points for lag 2

	Positive tail		Negative tail	
	5%	1%	5%	1%
$p = 5$ $(N = 10)$				
Exact	0.342	0.540	−0.417	−0.595
Approx.	0.360	0.525	−0.561	−0.705
$p = 7$ $(N = 14)$				
Exact	0.335	0.482	−0.479	−0.616
Approx.	0.335	0.485	−0.479	−0.615

7. Summary. 1. The exact and large sample distributions have been derived for the serial correlation coefficient for lag 1 and the exact significance points tabultaed for N, the number of observations, up to 75; for $N > 75$, the large sample approximations can be used.

2. It has been noted that the distributions for any lag L are the same as those for lag 1 when L and N are prime to each other. In general the distribution of the serial correlation coefficient can be derived for any L and N by using only those distributions for which L is a factor of N. The distributions and significance points have been derived for $N/L = p = 2, 3$ and 4. For $p > 4 (N > 4L)$, the significance points given for lag 1 probably can be used when L is greater than 4 or 5. The accuracy of this approximation has been checked for lag 2.

3. These significance points should be useful in determining the methods of studying a time series, as suggested by Wold, and in the formulation of a better test of the significance of regression coefficients when we know that the observations are correlated in time. In addition, we now have a method of testing our assumptions of independence for any set of data.

REFERENCES

[1] R. L. ANDERSON, *Serial Correlation in the Analysis of Time Series*, unpublished thesis, Library, Iowa State College, Ames, Iowa, 1941.

[2] M. S. BARTLETT, "Some aspects of the time-correlation problem in regard to tests of significance," *Roy. Stat. Soc. Jour.*, Vol. 98 (1935), pp. 536–543.

[3] W. G. Cochran, "Distribution of quadratic forms in a normal system with applications to the analysis of covariance," *Camb. Phil. Soc. Proc.*, Vol. 30 (1934), pp. 178–191.
[4] L. E. Dickson, *Modern Elementary Theory of Numbers*, U. of Chicago Press, 1939.
[5] Karl Pearson (Editor), *Tables of the Incomplete Beta-Function*, Cambridge U. Press, 1934.
[6] G. Tintner, "Tests of significance in time series," *Annals of Math. Stat.*, Vol. 10 (1939), p. 141 ff.
[7] G. Tintner, *The Variate Difference Method*, Principia Press, Bloomington, Indiana, Appendix 5B, 1940.
[8] H. Wold, *A Study in the Analysis of Stationary Time Series*, Almquist and Wiksells Boktryckeri A. B., Uppsala, 1938.
[9] G. U. Yule, "On the time-correlation problem," *Roy. Stat. Soc. Jour.*, Vol. 84 (1921), pp. 496–537.

SERIAL CORRELATION AND QUADRATIC FORMS IN NORMAL VARIABLES[1]

By Tjalling Koopmans

Penn Mutual Life Insurance Company

1. Estimation problems of stochastical processes. In regression analysis of economic time series a situation often arises in which a certain observed quantity represents a "dependent" variable at one time and an "independent" variable at a later time. For instance, the following relations may exist between the price x_t and the supply y_t of hogs at any time t:

$$x_t = \alpha - \beta y_t + z_t'$$
$$y_t = \gamma + \delta x_{t-1} + z_t''. \tag{1}$$

The first of these equations expresses the price-depressing influence of large supplies. The second equation expresses the supply-stimulating influence of high prices one time unit (in the case of hogs, about 18 months) earlier. The terms z_t' and z_t'' represent influences of additional variables and/or random disturbances. Elimination of y_t leads to

$$x_t = \epsilon - \zeta x_{t-1} + z_t . \tag{2}$$

The statistical estimation of the parameters ϵ and ζ of such an equation is usually attempted by the ordinary least squares method, disregarding the fact that the observation x_t is both a dependent variable at time t and an independent variable at time $t + 1$. The following simple example shows that this may lead to erroneous results particularly in small samples. Suppose that $\epsilon = 0$, $\zeta = -1$, and that z_t is a purely random variable with mean 0, while only three successive observations are available. The least squares estimate of η is then given by the slope of the straight line connecting the points (x_1 , x_2) and (x_2 , x_3) in the plane of x_{t-1} and x_t . This slope, however, has an expected value 0, because according to our assumptions the conditional expectation of x_3 for a prescribed value of x_2 is equal to x_2 , whatever value that is. Thus the least squares estimate of $\zeta = -1$ has an expected value 0 showing an important bias.

Mathematical business cycle theories utilize systems of equations much more complicated than the example considered [1]. The common feature of these equation systems is, however, that they reduce fluctuations in a set of economic variables to

1. earlier fluctuations in the same set of variables,
2. changes in given non-economic or external variables, and
3. random disturbances.

[1] This investigation was carried out at the Local and State Government Section (Princeton Surveys) of the School for Public and International Affairs of Princeton University. The main results were presented to the Chicago meeting of the Institute of Mathematical Statistics in September 1941.

An equation system of this type has been said to define a *stochastical process* in a number of variables [2]. The statistical testing of mathematical business cycle theories accordingly requires a theory of estimation of the parameters of stochastical processes. The operation of stochastical processes is also apparent in meteorological data. Assuming a normal distribution for the random disturbances, it will be seen that the mathematical prerequisite for an estimation theory of stochastical processes is the study of joint distributions of certain quadratic forms in normal variables.

In this article only the very simplest problem of this class will be treated, namely that of testing the significance of ζ in equation (2) if it is known that $|\zeta| < 1$ and that ϵ is equal to zero. This is the problem of testing the significance of single serial regression, or of single serial correlation, because the distinction between single regression and correlation coefficients disappears in this simple case for coefficients absolutely smaller than unity.

In the next section the problem of estimating single serial correlation if the mean is known will be stated and the difficulties involved will be discussed. In section 3 a conditional distribution of a quadratic form in normal variables will be derived. The proof in section 3 covers only forms in five or more variables, but another proof covering any number of variables is given in section 4. This distribution is then applied to devise a test of significance of serial correlation in section 5. The reading of section 4 is not necessary for the understanding of section 5. Readers desiring to locate only the main results can read those from equations (3), (11), (16), (21), (36), (61), (62), (74), (79), (82), (92), and (96).

2. The estimation of serial correlation. In the stochastical process

$$x_t = \rho x_{t-1} + z_t, \tag{3}$$

where the z_t are independent drawings from a normal distribution with mean 0 and standard deviation σ, the parameter ρ may have any positive or negative values. The process will only be a stationary one if

$$|\rho| < 1. \tag{4}$$

For, since

$$Ex_t = Ex_{t-1} = Ez_t = 0, \qquad Ez_t^2 = \sigma^2, \tag{5}$$

and

$$Ex_t^2 = \rho^2 Ex_{t-1}^2 + \sigma^2, \tag{6}$$

a variance of x_t independent of t will be possible only if (4) is satisfied, in which case

$$Ex_t^2 = \frac{\sigma^2}{1 - \rho^2}. \tag{7}$$

If (4) is not satisfied, however, Ex_t^2 will be an increasing function of t tending to infinity in approximately geometric progression if t increases indefinitely. In this article the limitation (4) will be imposed a priori.

It follows from (3), (7), and the assumption regarding z_1 that the joint distribution of the quantities $x_1, z_2, z_3, \cdots, z_T$ is given by

$$
(8) \qquad \left(\frac{1-\rho^2}{2\pi\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{1}{2}(1-\rho^2)x_1^2/\sigma^2} \cdot \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}(T-1)} e^{-\frac{1}{2}\sum_{t=2}^{T} z_t^2/\sigma^2} dx_1 dz_2 \cdots dz_T .
$$

Since the Jacobian of the transformation (3) from the variables $x_1, z_2, \cdots, z_T$ to the variables $x_1, x_2, \cdots, x_T$ equals unity, the joint distribution function of the T successive observations $x_1, x_2, \cdots, x_T$ that make up a sample is found simply by replacing the z_t in (8) by the corresponding expressions in the x's. This leads to the distribution

$$
(9) \qquad \frac{(1-\rho^2)^{\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{1}{2}T}} e^{-\frac{1}{2}\{l - 2\rho m + (1+\rho^2)n\}/\sigma^2} dx_1 dx_2 \cdots dx_T ,
$$

in which the three quadratic forms

$$
(10) \qquad \begin{aligned} l &= x_1^2 + x_T^2 , \\ m &= x_1x_2 + x_2x_3 + \cdots + x_{T-1}x_T , \\ n &= x_2^2 + x_3^2 + \cdots + x_{T-1}^2 , \end{aligned}
$$

are the only characteristics of the sample that enter. In other words, l, m and n are jointly sufficient statistics for the estimation of ρ and σ. It may be noted that these statistics remain the same if the series of observations is taken in inverse order.

It seems natural to attempt maximum likelihood estimation of ρ and σ, even if the usual optimal properties of estimates so obtained have so far not been proved for stochastical processes. Straightforward calculations lead to the following third-degree equation for the maximum likelihood estimate $\hat{\rho}$ of ρ:

$$
(11) \qquad (m - \hat{\rho}n)(1 - \hat{\rho}^2) - \frac{\hat{\rho}}{T}\{l - 2\hat{\rho}m + (1 + \hat{\rho}^2)n\} = 0.
$$

Of course the root asymptotically approaching m/n has to be selected. The corresponding maximum likelihood estimate $\hat{\sigma}$ of σ is given by

$$
(12) \qquad \hat{\sigma}^2 = \frac{1}{T}\{l - 2\hat{\rho}m + (1 + \hat{\rho}^2)n\}.
$$

In view of the complicated definition of $\hat{\rho}$ it seems desirable as a first step to derive from (9) the joint probability distribution of l, m and n. This requires a transformation of the volume element $dx_1 \cdots dx_T$ in (9) to the form

$$
(13) \qquad \phi(l, m, n)\, dl\, dm\, dn.
$$

which it assumes after integration over $T - 3$ other coordinates the variation of which does not change l, m and n.

Since this is purely a problem of integration completely defined by the expressions (10), the resulting function $\phi(l, m, n)$ is independent of ρ and σ. The joint distribution

$$(14) \qquad \frac{(1-\rho^2)^{\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{1}{2}T}} e^{-\frac{1}{2}(l - 2\rho m + (1+\rho^2)n)/\sigma^2} \phi(l, m, n)\, dl\, dm\, dn$$

of l, m and n will thus be known for any values of ρ and σ as soon as it is known for two particular values.

If as particular values we choose $\rho = 0$ and $\sigma = 1$, the x_t become identical with the z_t, and the problem is that of finding the joint distribution of the quadratic forms (10) in independent normal variables with mean 0 and variance 1. Even if so simplified, the problem is a complicated one. While there are infinitely many common sets of principal axis of the forms l and n, none of these sets of axis has a single axis in common with m.

Although no solution is offered for this problem, the following suggestion may be ventured. Once $\phi(l, m, n)$ is known, the mathematically simplest procedure for interval estimation of ρ might well be one that confines attention to samples having the same values of l and n as the sample actually obtained. Suitably chosen percentiles of the conditional distribution of m with l and n fixed at the observed values, would be convertible into confidence limits for ρ with the help of (14).

A simpler mathematical problem is encountered in testing whether the existence of a difference between ρ and 0 can be established, or, in other words, in testing the significance of serial correlation. If $\rho = 0$, the distribution function in (14) depends only on $p = l + n$, not on l or n separately, and exact significance limits for m can be derived from the joint distribution

$$(15) \qquad (2\pi\sigma^2)^{-\frac{1}{2}T} e^{-\frac{1}{2}p/\sigma^2} \psi(p, m)\, dp\, dm$$

of p and m only. This distribution will be studied in the next three sections. It is hoped that the methods there applied will provide a useful starting point in the treatment of other problems of the class described in section 1.

3. Distribution of a quadratic form in normal variables on the unit sphere. Consider two quadratic forms in T independent normal variables with mean 0 and variance 1,

$$(16) \qquad \begin{aligned} p &= x_1^2 + x_2^2 + \cdots + x_T^2, \\ q &= \kappa_1 x_1^2 + \kappa_2 x_2^2 + \cdots + \kappa_T x_T^2. \end{aligned}$$

While the characteristic values of the form p are all coincident with the value 1, the characteristic values κ_t of q are provisionally supposed to be different from each other, so that they can be arranged in decreasing order

$$\kappa_1 > \kappa_2 > \cdots > \kappa_T. \tag{17}$$

The probability density

$$(2\pi)^{-\frac{1}{2}T} e^{-\frac{1}{2}p} \tag{18}$$

in the space of the variables is constant on any sphere

$$p = p_0 = \text{constant}, \tag{19}$$

while the distribution function $g(p)$ of p is that of the χ^2-distribution with T degrees of freedom

$$g(p) = \frac{p^{\frac{1}{2}T-1} e^{-\frac{1}{2}p}}{2^{\frac{1}{2}T}\,\Gamma(\frac{1}{2}T)}. \tag{20}$$

The hyper-surfaces on which the ratio

$$r = \frac{q}{p} \tag{21}$$

of q to p is constant are cones with the origin as vertex dissecting the same proportion of the metric "surface" of each sphere (19). It follows that the conditional distribution function of r for a prescribed value p_0 of p is independent of that value p_0, and is therefore equal to the unrestricted distribution function $h(r)$ of r. In other words, p and r are independently distributed. Their joint distribution being

$$g(p)h(r)\,dp\,dr, \tag{22}$$

the joint distribution of p and $q = rp$ is found to be

$$f(p, q)\,dp\,dq = g(p)h\left(\frac{q}{p}\right)dp\,\frac{dq}{p} = \frac{g(p)}{p}\,h\left(\frac{q}{p}\right)dp\,dq. \tag{23}$$

The function $h(\ )$ may therefore also be described as the conditional distribution function of q on the unit sphere

$$p = 1. \tag{24}$$

Since κ_1 and κ_T are the extreme values of q under the condition (24), the function $h(r)$ vanishes outside these limits.

We shall now derive an expression for $h(r)$ by comparing (23) with an expression for $f(p, q)$ obtained through the inversion theorem of characteristic functions. The characteristic function $F(\eta, \theta)$ corresponding to the variables p and q is

$$F(\eta, \theta) = (2\pi)^{-\frac{1}{2}T} \int e^{-\frac{1}{2}p + i(\eta p + \theta q)}\,dx_1 \cdots dx_T = D^{-\frac{1}{2}}(\eta, \theta), \tag{25}$$

where, according to (16), the polynomial $D(\eta, \theta)$ is given by

$$D(\eta, \theta) = \prod_{t=1}^{T} (1 - 2i\eta - 2i\theta\kappa_t). \tag{26}$$

It follows from the inversion theorem that

$$(27) \qquad f(p, q) = (2\pi)^{-2} \iint\limits_{-\infty}^{\infty} e^{-i(\eta p+\theta q)} D^{-1}(\eta, \theta)\, d\eta\, d\theta,$$

the order of integration over η and θ being immaterial.

Any elementary factor of $D(\eta, \theta)$ may be written

$$(28) \qquad d_t(\eta, \theta) = 1 - 2i\eta - 2i\theta\kappa_t = (1 - 2i\eta)\left(1 - \frac{2i\theta\kappa_t}{1 - 2i\eta}\right).$$

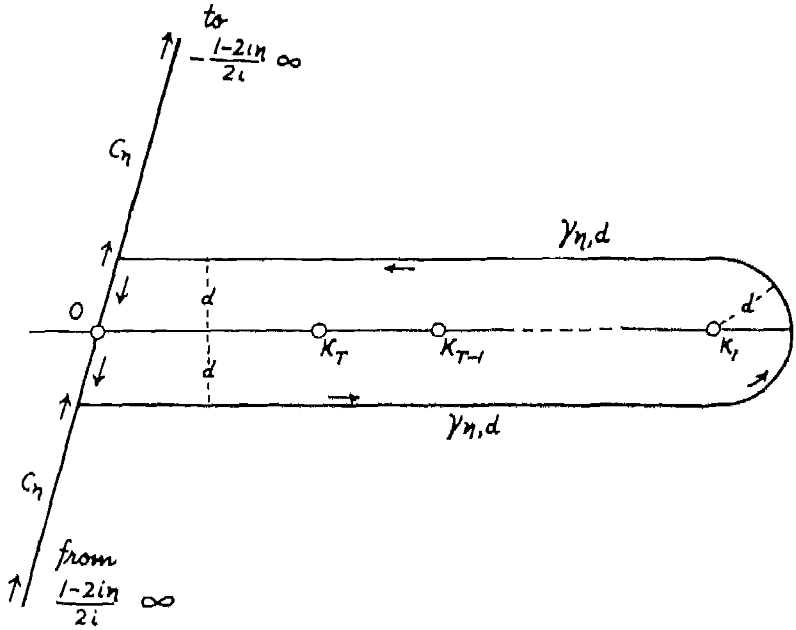


FIGURE 1. Paths of integration in the κ-plane

First considering the integration over θ (while η has some fixed value), we may instead of θ use

$$(29) \qquad \kappa = \frac{1 - 2i\eta}{2i\theta}$$

as an integration variable. The path of integration c_η in the κ-plane then is a straight line from 0 to $-\dfrac{1 - 2i\eta}{2i}\infty$ and another straight line from $\dfrac{1 - 2i\eta}{2i}\infty$ back to 0, as indicated in Figure 1, and the transformed integral (27) runs

$$f(p, q) = (2\pi)^{-2} \int_{-\infty}^{\infty} \left[e^{-i\eta p}(1 - 2i\eta)^{-\frac{1}{2}T+1} \int_{c_\eta} e^{-(1-2i\eta)q/2\kappa} \cdot \left\{ \prod_{t=1}^{T} \left(1 - \frac{\kappa_t}{\kappa}\right) \right\}^{-1} \left(-\frac{1}{2i\kappa^2}\right) d\kappa \right] d\eta.$$

The integrand

$$e^{-(1-2i\eta)q/2\kappa}\prod_t\left(1-\frac{\kappa_t}{\kappa}\right)\left(-\frac{1}{2i\kappa}\right) \tag{31}$$

for the integration over κ has singularities only in the points $\kappa = 0$ and $\kappa = \kappa_t$, $t = 1, 2 \cdots T$. In order to simplify the argument we shall suppose that the quadratic form q is positive definite, or, in connection with (17), that $\kappa_T > 0$. The location of the singularities is then as pictured in Figure 1. As $\kappa \to \infty$ the integrand (31) is regular and of the order of magnitude of κ. Consequently a curve integral of (31) along the whole or any part of the circle $|\kappa| = R$ will tend to 0 if R tends to infinity. Using a theorem of Cauchy, it is therefore permissible in (30) to replace the described path c_η by another path $c_{\eta,R}$ which starts out along c_η from 0 up to $-\frac{1-2i\eta}{2i}R$, from there follows the circle $|\kappa| = R$ to the right over an angle π up to the point $\frac{1-2i\eta}{2i}R$, and from there returns to 0 along c_η -provided that $R > \kappa_1$. After reversing the direction in which the path is followed in order to do away with the negative sign in (30), the path so obtained can again be replaced by the path $\gamma_{\eta,d}$ shown in Figure 1, which coincides with c_η only up to a small distance d from the real axis, and encircles all singularities κ_t while retaining a distance d from the part of the real axis to the left of and up to κ_1. Finally, a path of integration γ' independent of the value of η is obtained by going to the limit in which $d \to 0$. This means integration twice along the part of the real axis between 0 and κ_1, integrating from 0 to κ_1 that branch of the integrand which is obtained by passing "under" each singularity, and going back from κ_1 to 0 with the branch obtained by passing "around" κ_1 and "over" each other singularity[2]. The integral so obtained converges at each singularity. This is also true for the singularity $\kappa = 0$ because we are dealing only with positive values of q, which makes the exponential factor in (31) tend to 0 if κ approaches zero. We shall now show that if in (30) the path γ' is substituted for c_η (with a change in sign), the order of integration over κ and η can be reversed if $T \geqq 5$.

The integral over η, taken from (30),

$$I = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-i\eta(p-q/\kappa)}(1-2i\eta)^{-\frac{1}{2}T+1}\,d\eta, \tag{32}$$

(in which κ is now a positive real number), is by the substitution $\chi^2 = p - q/\kappa$ transformed to the integral encountered in the derivation of the χ^2-distribution (with $T - 2$ degrees of freedom) by the inversion theorem of characteristic functions. It may be quoted without proof (see [3] p. 42) that it equals

$$\begin{cases} I = \dfrac{(p-q/\kappa)^{\frac{1}{2}T-2}e^{-\frac{1}{2}(p-q/\kappa)}}{2^{\frac{1}{2}T-1}\Gamma(\frac{1}{2}T-1)}, & \text{if } p - q/\kappa \geqq 0, \text{ or } \kappa \geqq r, \\ I = 0, & \text{if } p - q/\kappa \leqq 0, \text{ or } \kappa \leqq r. \end{cases} \tag{33}$$

[2] For even values of T the parts of γ' for which $\kappa < \kappa_T$ can be disregarded, because on these parts the same branch of the integrand is integrated in opposite directions.

It is necessary to observe, however, that the integral I converges uniformly for all real values of κ whenever $T \geqq 5$, because then

$$(34) \qquad \int_{-\infty}^{\infty} |\, 1 - 2i\eta \,|^{-\frac{1}{2}T+1} \, d\eta,$$

is convergent. Because of this property, the reversal of the order of integration is allowed for $T \geqq 5$.

If now in (30) we first carry out the integration over η and use (33), we are left with

$$(35) \qquad f(p, q) = \frac{e^{-\frac{1}{2}p}}{2^{\frac{1}{2}T}\Gamma(\frac{1}{2}T - 1)} \int_{\gamma_r} \left(p - \frac{q}{\kappa}\right)^{\frac{1}{2}T-2} \left\{\prod_l \left(1 - \frac{\kappa_l}{\kappa}\right)\right\}^{-\frac{1}{2}} \frac{d\kappa}{2i\kappa^2},$$

where γ_r now is any curve proceeding from $\kappa = r$ into the lower half-plane, crossing the real axis at a point $\kappa > \kappa_1$, and returning to $\kappa = r$ through the upper half-plane, as indicated in Figure 2. (The path directly obtained is a path γ_r' consisting of twice the real axis between r and κ_1, the branches of the integrand being taken as indicated by γ_r.) Comparing (35) with (23) and (20), using (21)

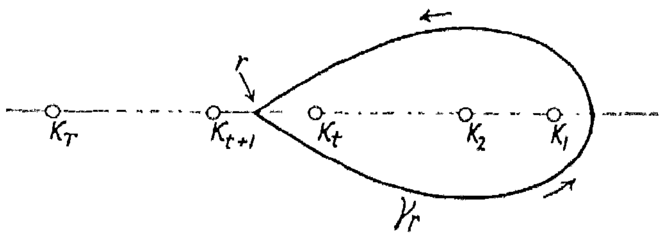


FIGURE 2. The integration path γ_r

and the well-known formula $\Gamma(x) = (x - 1)\Gamma(x - 1)$, we find the following expression for the distribution function of r:

$$(36) \qquad h(r) = \frac{\frac{1}{2}T - 1}{2\pi i} \int_{\gamma_r} \frac{(\kappa - r)^{\frac{1}{2}T-2}}{\prod_l (\kappa - \kappa_l)^{\frac{1}{2}}} \, d\kappa.$$

This function vanishes for $r \geqq \kappa_1$. In order to arrive at a positive distribution function for $\kappa_T < r < \kappa_1$ that branch of the integrand must be selected which is positive for real values of κ exceeding κ_1.

It is worth noting that the degree in κ of the numerator of the integrand is two less than that of the denominator. Owing to this fact, indeed, the distribution function $h(r)$ satisfies the two obvious conditions:

$$(37) \qquad h(r) = 0 \quad \text{for} \quad r \leqq \kappa_T, \qquad \int_{\kappa_T}^{\kappa_1} h(r) \, dr = 1.$$

For $r \leqq \kappa_T$ the path of integration in (36) can be replaced by any closed contour enclosing all the singularities $r, \kappa_T, \cdots, \kappa_1$ (r is a singularity only if T is odd). Taking as such a contour the circle $|\, \kappa \,| = R$ with R tending to infinity, we find that $h(r) = 0$ because the integrand is of an order κ^{-2} at $\kappa = \infty$. Further, if γ_r is again replaced by γ_r' which runs entirely along the real axis,

$$(38)\quad \begin{cases} \displaystyle\int_{\kappa_T}^{\kappa_1} h(r)\,dr = \frac{\frac{1}{2}T-1}{2\pi i}\int_{\kappa_T}^{\kappa_1}\left[\int_{\gamma'_r}\frac{(\kappa-r)^{\frac{1}{2}T-2}}{\prod_t(\kappa-\kappa_t)^{\frac{1}{2}}}\,d\kappa\right]dr \\[2ex] \displaystyle\qquad = \frac{\frac{1}{2}T-1}{2\pi i}\int_{\gamma'_{\kappa_T}}\left[\prod_t(\kappa-\kappa_t)^{-\frac{1}{2}}\int_{\kappa_T}^{\kappa}(\kappa-r)^{\frac{1}{2}T-2}\,dr\right]d\kappa \\[2ex] \displaystyle\qquad = \frac{1}{2\pi i}\int_{\gamma'_{\kappa_T}}\frac{(\kappa-r)^{\frac{1}{2}T-1}}{\prod_t(\kappa-\kappa_t)^{\frac{1}{2}}}\,d\kappa = 1, \end{cases}$$

because the integrand in the last integral is of the order of κ^{-1} at the point $\kappa = \infty$.

The quantities r and κ_t enter into the right hand member of (36) only in the form of differences from the integration variable κ. The addition of a constant ϵ to both r and the κ_t will therefore merely result in a change of location of the distribution on the r-axis without a change in form:

$$(39)\qquad h^*(r+\epsilon) = h(r).$$

This could be expected since such a transformation means the addition of ϵp to the quadratic form q studied. It follows that the validity of (36) is not limited to positive definite quadratic forms q, since any other quadratic form can be transformed to a positive definite form by this operation if a sufficiently large value of ϵ is taken.

The function $h(r)$ is a different analytic function between any two different successive characteristic values κ_t and κ_{t+1}. The expression (36) holds for even and for odd values of T, and is also valid for any number of coincidences in the set of characteristic values κ_t. It is true that integration along the paths γ' or γ'_r entirely coincident with the real axis, such as has been introduced in intermediate stages of the above proof, cannot be done if two or more of the κ_t coincide, because of divergence of the integral. Once (36) has been established for distinct characteristic values, however, it follows from considerations of continuity that this result holds good also if coincidences occur in the set κ_t.

The function $h(r)$ has been studied by von Neumann [4] by an entirely different and very ingeneous method for the special case that T is even while no two characteristic values are equal, and for the case that the characteristic values are equal two by two but otherwise different. The properties established by von Neumann, and some generalizations of these properties, can be derived from (36). If T is even, the derivative of $h(r)$ of order $\frac{1}{2}T-1$ is

$$(39)\quad \left(\frac{d}{dr}\right)^{\frac{1}{2}T-1} h(r) \begin{cases} \displaystyle = \frac{(\frac{1}{2}T-1)!}{\pi}\cdot\frac{(-1)^{\frac{1}{2}(T-t+1)}}{\prod_{s=1}^{T}|r-\kappa_s|^{\frac{1}{2}}} & \text{if } \kappa_{t+1} < r < \kappa_t \text{ and } t \text{ odd,} \\[2ex] \text{does not exist for } r = \kappa_t,\ t = 1, 2 \cdots T, & \\[1ex] = 0 \quad \text{for all other values of } r. & \end{cases}$$

If all characteristic values are distinct, all derivatives of an order lower than $\frac{1}{2}T - 1$ exist and are continuous everywhere. Generally, whether T be even or odd, at a point where k characteristic values coincide $\left(\frac{d}{dr}\right)^j h(r)$ will exist and will be continuous if $j \leqq \frac{1}{2}(T - k) - \frac{3}{2}$, and will not exist if $j \geqq \frac{1}{2}(T - k) - 1$.

If the characteristic values are pairwise equal,

$$\kappa_{2s-1} = \kappa_{2s} = \lambda_s, \qquad s = 1, 2, \cdots S, \tag{40}$$

but otherwise distinct, their total number $T = 2S$ must be even, and the only singularities of the integrand in (36) are poles at the points $\kappa = \lambda_s$. Accordingly the path of integration γ_r can be considered as a closed curve, and the integral in (36) can be replaced by the sum of the residuals of the integrand at all poles inside the curve:

$$h(r) = (S - 1) \sum_{s=1}^{s_r} \frac{(\lambda_s - r)^{S-2}}{P'(\lambda_s)}, \quad \text{if } \lambda_{s_r+1} < r < \lambda_{s_r}. \tag{41}$$

Here $P'(\lambda)$ is the derivative of

$$P(\lambda) = \prod_{s=1}^{S} (\lambda - \lambda_s), \tag{42}$$

its value in the point $\lambda = \lambda_s$ being

$$P'(\lambda_s) = \left[\frac{P(\lambda)}{\lambda - \lambda_s}\right]_{\lambda=\lambda_s} = \prod_{\substack{u=1 \\ (u \neq s)}}^{S} (\lambda_s - \lambda_u). \tag{43}$$

For $S = 2$ this is simply the rectangular distribution

$$h(r) = \frac{1}{\lambda_1 - \lambda_2}, \qquad \lambda_2 < r < \lambda_1. \tag{44}$$

The numerical calculation of the distribution (36) with distinct characteristic values is extremely cumbersome except for very small values of T. If the characteristic values κ_t follow some definite pattern, however, it may be possible in some instances to work out a reasonable approximation formula. Two examples of this type will be shown in section 5.

4. **Another proof that covers also cases with $T < 5$.** The proof of (36) given above holds only for $T \geqq 5$. Once the form of (36) is known or presumed, however, another proof of its validity is available, which has mathematical interest in itself, and covers all cases from $T = 2$ upwards. This is a proof by complete induction, based on the proposition that, if (36) holds for T variables, then it also holds for $T + 1$ variables. This proposition again rests on the recurrent relation

$$h_{T+1}(r') = \frac{\Gamma(\frac{1}{2}T + \frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}T)} (r' - \kappa_{T+1})^{\frac{1}{2}T-1} \int_{r'}^{\kappa_1} (r - r')^{-\frac{1}{2}}(r - \kappa_{T+1})^{-\frac{1}{2}T+\frac{1}{2}} h_T(r)\, dr, \tag{45}$$

$$\text{if } \kappa_{T+1} < r' < \kappa_1 \text{ and } \kappa_{T+1} < \kappa_T,$$

proved elsewhere in this issue by von Neumann[3] [5]. It connects the distribution function $h_T(r)$ for T variables with the function $h_{T+1}(r')$ obtained by the addition of one variable x_{T+1} and one characteristic value κ_{T+1}.

We shall substitute the "presumed" expression (36) for $h_T(r)$ with $T \geq 3$ in (45) in order to show that the result for $h_{T+1}(r)$ is the same expression with T increased by one. In this proof it has for simplicity's sake been assumed that the new characteristic value κ_{T+1} is smaller than any of those already present, and that no two of the κ_t are equal. It is then possible again to select in (36) the path of integration γ_r' which proceeds along the real axis from r to κ_1 and returns along the real axis to r, passing each singularity in the same way as γ_r does. If the integral (36) is substituted in (45) in this form, the order of integration over κ and r can be reversed, the result being

$$\begin{aligned} h_{T+1}(r') &= \frac{\frac{1}{2}T - 1}{2\pi i} \frac{\Gamma(\frac{1}{2}T + \frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}T)} (r' - \kappa_{T+1})^{\frac{1}{2}T-1} \\ &\cdot \int_{\gamma_{r'}'} \left[\left\{ \prod_{t=1}^{T} (\kappa - \kappa_t) \right\}^{-\frac{1}{2}} \int_{r'}^{\kappa} (r - r')^{-\frac{1}{2}}(r - \kappa_{T+1})^{-\frac{1}{2}T+\frac{1}{2}}(\kappa - r)^{\frac{1}{2}T-1} dr \right] d\kappa \end{aligned} \tag{46}$$

Writing for greater clarity $\kappa_{T+1} = a$, $r' = b$, $\kappa = c$, $r = z$, we have to evaluate the integral

$$I_T(a, b, c) = \int_b^c (z - a)^{-\frac{1}{2}T+\frac{1}{2}}(z - b)^{-\frac{1}{2}}(c - z)^{\frac{1}{2}T-1} dz, \tag{47}$$

$$a < b < c, \quad T \geq 3,$$

with the positive square roots taken if z is real and $b < z < c$. Suppose first that $T = 2S + 1$ or odd. Then the integrand

$$\phi_{2S+1}(z) = (z - a)^{-S}(z - b)^{-\frac{1}{2}}(c - z)^{S-\frac{1}{2}} \tag{48}$$

has singularities at a, b and c, of which only those at b and c are of a type such that $\phi_{2S+1}(z)$ changes its sign if the argument z is turned once around the singularity. It follows that

$$2I_{2S+1} = \int_\delta \phi_{2S+1}(z)\, dz, \tag{49}$$

the path of integration δ being as indicated in Figure 3. For if the curve δ is contracted so as to run entirely along the real axis, from b to c and back to b, the two parts of the curve will each yield a contribution equal to I_{2S+1}, the understanding being that positive square roots are taken when going from b to c.

The integrand $\phi_{2S+1}(z)$ is regular at $z = \infty$ and of order z^{-1} in a neighborhood of that point. It follows that

---

[3] I am greatly indebted to Professor von Neumann for communicating this relation to me before its publication.

$$-2I_{2S+1} = \int_\epsilon \phi_{2S+1}(z)\,dz, \tag{50}$$

where ϵ, as in Figure 3, encloses the only singularity not enclosed by δ. In a neighborhood of $z = a$ the following expansion of $\phi_{2S+1}(z)$ holds:

$$\phi_{2S+1}(z) = \sum_{s=0}^{\infty} \frac{(z-a)^{-S+s}}{s!}\left[\left(\frac{\partial}{\partial z}\right)^s (z-b)^{-\frac12}(c-z)^{S-\frac12}\right]_{z=a}. \tag{51}$$

The only term contributing to (50) is that with $-S + s = -1$. Since we selected a branch of $\phi_{2S+1}(z)$ such that $(z-b)^{-\frac12}(c-z)^{S-\frac12}$ falls on the positive pure imaginary axis for real values of z below b, this term can be written

$$(z-a)^{-1}\frac{i}{(S-1)!}\left(\frac{\partial}{\partial a}\right)^{S-1}(b-a)^{-\frac12}(c-a)^{S-\frac12}, \tag{52}$$

where positive square roots should now be taken. The contribution of this term in (50) is $2\pi i$ times the coefficient of $(z-a)^{-1}$, and therefore

$$I_{2S+1} = \frac{\pi}{(S-1)!}\left(\frac{\partial}{\partial a}\right)^{S-1}(b-a)^{-\frac12}(c-a)^{S-\frac12}$$

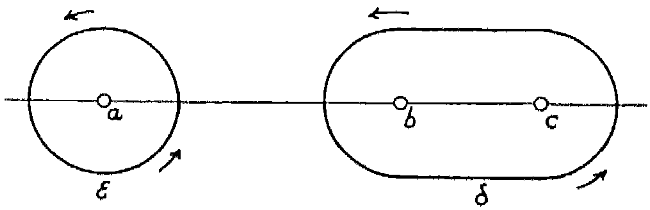


FIGURE 3. The integration paths δ and ϵ

$$= \frac{\frac12\cdot\frac32\cdot\,\cdots\,(S-\frac32)\pi}{(S-1)!}(-1)^{S-1}\sum_{s=0}^{S-1}\binom{S-1}{s}(-1)^s(b-a)^{-\frac12-s}(c-a)^{-\frac12+s} \tag{53}$$

$$= \frac{\Gamma(S-\frac12)\pi}{\Gamma(\frac12)\Gamma(S)}(b-a)^{-S+\frac12}(c-a)^{-\frac12}(c-b)^{S-1}.$$

Since $\Gamma(\frac12) = \pi^{\frac12}$, it follows that for odd values of T

$$I_T = \frac{\Gamma(\frac12T-1)\Gamma(\frac12)}{\Gamma(\frac12T-\frac12)}(b-a)^{-\frac12T+1}(c-a)^{-\frac12}(c-b)^{\frac12(T-3)}. \tag{54}$$

It is easily seen that the same relation holds good if $T = 2S$ is éven. In that case it follows from (47) that

$$\begin{aligned}\left(\frac{\partial}{\partial c}\right)^{S-1} I_{2S} &= \frac{\partial}{\partial c}(S-2)!\int_b^c (z-a)^{-S+\frac12}(z-b)^{-\frac12}\,dz \\ &= (S-2)!\,(c-a)^{-S+\frac12}(c-b)^{-\frac12}.\end{aligned} \tag{55}$$

In a manner similar to the transformations in (53) it can likewise be proved that the right hand member in (54) has the same derivative of order $S-1$

with respect to c. It follows that the two members of (54) differ by a polynomial $Q(c)$ in c of a degree at most equal to $S - 2$, the coefficients of which may depend on a and b. However, both members of (54) as well as their first $S - 2$ derivatives with respect to c vanish if $c = b$. Therefore $Q(c)$ vanishes identically, and (54) holds for any integral values of T not smaller than 3.

Finally, if (54) is inserted in (46) an expression for $h_{T+1}(r')$ is obtained which corresponds to (36) with T replaced by $T + 1$.

It remains to prove (36) for some initial value of T. For $T = 2$ the integral in (36) is divergent, but the form of $h(r)$ is easily found directly. Writing

$$p = x_1^2 + x_2^2, \qquad r = \frac{q}{p} = \frac{\kappa_1 x_1^2 + \kappa_2 x_2^2}{x_1^2 + x_2^2}, \tag{56}$$

we find that

$$\begin{aligned} \frac{\partial(x_1, x_2)}{\partial(p, r)} &= \left[\frac{\partial(p, r)}{\partial(x_1, x_2)}\right]^{-1} = \begin{vmatrix} 2x_1 & 2x_2 \\ \frac{2x_1}{p}(\kappa_1 - r) & \frac{2x_2}{p}(\kappa_2 - r) \end{vmatrix}^{-1} \\ &= -\frac{p}{4x_1 x_2(\kappa_1 - \kappa_2)} = -\frac{1}{4(\kappa_1 - r)^{\frac{1}{2}}(r - \kappa_2)^{\frac{1}{2}}}. \end{aligned} \tag{57}$$

The probability density in the x_1-x_2-plane is, of course, $(2\pi)^{-1}e^{-\frac{1}{2}p}$, but in making the transformation (57) a factor 4 must be applied to account for the fact that to given values of p and r correspond 4 sets of values of x_1 and x_2, differing in the signs only. This leads to the joint distribution of p and r

$$\frac{1}{2\pi} e^{-\frac{1}{2}p} \frac{dp\, dr}{(\kappa_1 - r)^{\frac{1}{2}}(r - \kappa_2)^{\frac{1}{2}}}, \tag{58}$$

and, after integration over p, to

$$\begin{aligned} h_2(r) &= \frac{1}{\pi(\kappa_1 - r)^{\frac{1}{2}}(r - \kappa_2)^{\frac{1}{2}}}, \quad \text{if } \kappa_2 < r < \kappa_1, \\ &= 0, \quad \text{if } r < \kappa_2 \quad \text{or} \quad \kappa_1 < r, \end{aligned} \tag{59}$$

in accordance with (39).

Finally, if (59) is inserted in (45) with $T = 2$, the result is

$$h_3(r') = \frac{1}{2\pi} \int_{[\kappa_3, r']}^{\kappa_1} \frac{(r - r')^{-\frac{1}{2}}}{(\kappa_1 - r)^{\frac{1}{2}}(r - \kappa_2)^{\frac{1}{2}}(r - \kappa_3)^{\frac{1}{2}}}\, dr, \qquad \kappa_3 < r' \leq \kappa_1, \tag{60}$$

if $[\kappa_2, r']$ denotes the largest of κ_3 and r'. Writing κ for r, we find that this integral is equivalent to that in (36) for $T = 3$, taking into account the rule established for selecting the branch of the integrand in (36). For, taking the path of integration γ_r' coincident with the real axis, the equal contributions from the two parts of the path between κ_2 and κ_1 reinforce each other, while for $r' < \kappa_2$ the remaining contributions (intervals between r' and κ_3) add up to zero. This completes the second proof of (36).

5. **Application to serial correlation.** We shall now derive the characteristic values κ_t in the case that

$$q = m = x_1x_2 + x_2x_3 + \cdots + x_{T-1}x_T. \tag{61}$$

It will be of interest to compare this case with the slightly modified case of the quadratic form

$$\bar{m} = x_1x_2 + x_2x_3 + \cdots + x_{T-1}x_T + x_Tx_1, \tag{62}$$

which contains an additional term x_Tx_1 accomplishing a circular arrangement of the variables This modification was originally suggested by Hotelling in order to simplify the characteristic polynomial. Other simplifications arising out of the circular arrangement will appear below. It is possible, of course, that the power of the test of significance of serial correlation is slightly affected by the substitution of $\bar{m}$ for m, but this presumption needs corroboration by a study of power functions.

The characteristic values of m are those values of κ for which the determinant of order T

$$\Delta_T = \begin{vmatrix} -\kappa & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{2} & -\kappa & \frac{1}{2} & \cdots & 0 \\ 0 & \frac{1}{2} & -\kappa & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & & -\kappa \end{vmatrix} = 0. \tag{63}$$

By development according to elements of the first row we find that

$$\Delta_T = -\kappa\Delta_{T-1} - \tfrac{1}{4}\Delta_{T-2}, \tag{64}$$

from which it follows that

$$\Delta_T = c_1\xi_1^T + c_2\xi_2^T, \tag{65}$$

if ξ_1 and ξ_2 are the roots of

$$\xi^2 + \kappa\xi + \tfrac{1}{4} = 0, \tag{66}$$

satisfying

$$\xi_1 + \xi_2 = -\kappa, \qquad \xi_1\xi_2 = \tfrac{1}{4}. \tag{67}$$

By inserting the known values of Δ_1 and Δ_2 in (65), the values of c_1 and c_2 are easily found to be such that

$$\Delta_T = \frac{\xi_1^{T+1} - \xi_2^{T+1}}{\xi_1 - \xi_2}. \tag{68}$$

Although as a polynomial in κ this is a rather complicated expression, the implicit form (68) will suffice for finding the roots of (63). Expressing all other variables in terms of one new variable ω,

$$\xi_1 = -\frac{\omega}{2}, \qquad \xi_2 = -\frac{1}{2\omega}, \qquad \kappa = \tfrac{1}{2}(\omega + \omega^{-1}), \tag{69}$$

we find for (68),

$$\Delta_T = (-\tfrac{1}{2})^T \frac{\omega^{T+1} - \omega^{-T-1}}{\omega - \omega^{-1}} = \left(-\frac{1}{2\omega}\right)^T \frac{\omega^{2(T+1)} - 1}{\omega^2 - 1}. \tag{70}$$

The only values of ω for which this expression vanishes are the roots of

$$\omega^{2(T+1)} = 1, \tag{71}$$

excepting those that are also roots of

$$\omega^2 = 1. \tag{72}$$

This leaves us with

$$\omega = e^{\pm \pi i t/(T+1)}, \qquad t = 1, 2, \cdots T. \tag{73}$$

The corresponding characteristic values are

$$\kappa_t = \cos \frac{\pi t}{T+1}, \qquad t = 1, 2 \cdots T, \tag{74}$$

because the same value of κ_t is obtained whether the positive or the negative sign is taken in (73). These are T different values κ_t, and hence each one is a single root of (63).

The characteristic values of $\bar{m}$ can now be derived from (68), although a simple straightforward method based on the properties of circulants is also available (see [6], p. 13). Writing

$$\bar{\Delta}_T = \begin{vmatrix} -\kappa & \frac{1}{2} & 0 & \cdots & \frac{1}{2} \\ \frac{1}{2} & -\kappa & \frac{1}{2} & \cdots & 0 \\ 0 & \frac{1}{2} & -\kappa & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & 0 & 0 & \cdots & -\kappa \end{vmatrix} = \Delta_T + 2(-1)^{T-1}(\tfrac{1}{2})^T - \tfrac{1}{4}\Delta_{T-2}, \tag{75}$$

we find easily from (70) that

$$\begin{aligned} \bar{\Delta}_T &= (-\tfrac{1}{2})^T \left(\frac{\omega^{T+1} - \omega^{-T+1}}{\omega - \omega^{-1}} - 2 - \frac{\omega^{T-1} - \omega^{-T+1}}{\omega - \omega^{-1}} \right) \\ &= (-\tfrac{1}{2})^T(\omega^T + \omega^{-T} - 2) = -(-\tfrac{1}{2})^{T-1}(\cos T\alpha - 1), \end{aligned} \tag{76}$$

if

$$\omega = e^{i\alpha}. \tag{77}$$

A complete set of the values ω_t for which $\bar{\Delta}_T$ vanishes is found from

$$\alpha_t = \frac{2\pi t}{T}, \qquad t = 1, 2 \cdots T, \tag{78}$$

and the corresponding characteristic values[4] $\bar{\kappa}_t$ are, according to (69),

[4] In order to simplify the formulae, the numbering of characteristic values according to decreasing size has been abandoned in (79).

$$(79) \qquad \bar{\kappa}_t = \cos \alpha_t = \cos \frac{2\pi t}{T}, \qquad t = 1, 2 \cdots T.$$

In contradistinction to the case without circular arrangement, the characteristic values with indices t and $T - t$ now coincide, such that all characteristic values are double except one ($\bar{\kappa}_T = 1$) if T is odd, and except two ($\bar{\kappa}_T = 1$, $\bar{\kappa}_{\frac{1}{2}T} = -1$) if T is even.

Taking advantage of the duplicity of almost all characteristic values, Anderson [6] has derived expressions equivalent to (36) for this case, using methods that depend on this particular condition. On the basis of these results he has computed 99- and 95-percentiles in the distribution of $\bar{r} = \frac{\bar{m}}{p}$ for the values $T = 2, 3, 4, 5, 6, 7, 9, 11, 13, 15, 25, 45$, interpolating the percentiles for intermediate values of T. The 95-percentile for $T = 45$ is 0 240, as compared with 0.261 for the normal distribution that provides an asymptotic approximation.

Whereas on this showing the normal approximation is slow in becoming accurate with increasing T, a method for obtaining a much closer approximation is available, which works out simplest with respect to $\bar{r}$, but can also be applied to r. The principle of this method is applicable whenever the characteristic values follow a definite mathematical pattern.

The method consists in replacing the finite number of discrete values $\bar{\kappa}_t$ in (36) by a continuous variable λ, distributed according to a density function suggested by, and as closely as possible approximating to, the scatter of the values $\bar{\kappa}_t$. According to (79) the values $\bar{\kappa}_t$ are ordinates of the cosine function at equidistant points spaced out so as to cover one complete period 2π of that function. It is natural to approximate this scatter by the density function

$$(80) \qquad \bar{\chi}(\lambda) = \frac{T}{\pi(1 - \lambda^2)^{\frac{1}{2}}},$$

of the cosine $\lambda = \cos \frac{2\pi\bar{\tau}}{T}$ of an expression in which the variable $\bar{\tau}$ has a rectangular distribution between 0 and T. The numerical factor in (80) is such that

$$(81) \qquad \int_{-1}^{1} \bar{\chi}(\lambda)\, d\lambda = T$$

equals the total number of characteristic values to be replaced by a density function. The idea underlying the substitution of $\bar{\chi}(\lambda)$ for the $\bar{\kappa}_t$ is to obtain what intuitively seems to be in some sense the closest approximation to the exact distribution function $\bar{h}(\bar{r})$ that has continuous derivatives of any order in any point except the two points ($\bar{r} = -1$ and $\bar{r} = +1$) that limit its range.

The factor in the integrand in (36) which involves the $\bar{\kappa}_t$ is approximated as follows:

$$(82) \qquad \prod_{t=1}^{T} (\kappa - \bar{\kappa}_t)^{-\frac{1}{2}} = e^{-\frac{1}{2}\sum_{t=1}^{T} \log(\kappa - \bar{\kappa}_t)} \sim \exp\left[-\frac{T}{2\pi}\int_{-1}^{1} \frac{\log(\kappa - \lambda)}{(1 - \lambda^2)^{\frac{1}{2}}}\, d\lambda\right].$$

In order to evaluate the integral

$$J = \int_{-1}^{1} \frac{\log(\kappa - \lambda)}{(1 - \lambda^2)^{\frac{1}{2}}} \, d\lambda \tag{83}$$

we shall first prove that its real part is independent of κ, or that

$$\Re \int_{-1}^{1} \frac{\log(\kappa - \lambda) - \log(-\lambda)}{(1 - \lambda^2)^{\frac{1}{2}}} \, d\lambda = 0, \tag{84}$$

if $\Re$ denotes "the real part of". The integrand in (84) has singularities at the points $\lambda = -1, 0, \kappa, 1$. These are of two types. The singularities $\lambda = \pm 1$ are introduced by the denominator and make the integrand change its sign if the argument λ is turned once around either singularity. If starting from a point on the real axis we turn the argument λ once around either of the other singularities, $\lambda = 0$ and $\lambda = \kappa$, introduced by the numerator, then the real part of the integrand is not affected, while $2\pi i$ or $-2\pi i$ is added to the imaginary part of the numerator, depending on the sense (clockwise or anti-clockwise) of the rotation *and* on the sign of the logarithm in (84) responsible for the singular-

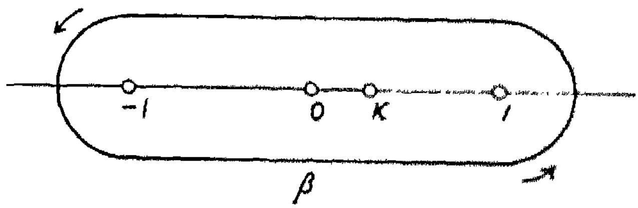


FIGURE 4. The integration path β

ity. It follows that one revolution along a closed curve β containing all four of the singularities, as indicated in Figure 4, carries us back to the same branch of the integrand, after mutually offsetting additions to the imaginary part of the numerator and after two changes in sign. This is in accordance with the regular character of the integrand at the point $\lambda = \infty$.

It follows furthermore that the left hand member of (84) can be replaced by

$$\tfrac{1}{2}\Re \int_{\beta} \frac{\log(\kappa - \lambda) - \log(-\lambda)}{(1 - \lambda^2)^{\frac{1}{2}}} \, d\lambda. \tag{85}$$

For, if the curve β is constricted to a path β' running along the real axis from -1 to $+1$ and back to -1, the contributions of the two halves of the path will be equal to each other, also with respect to sign. This is also true for the parts of the path β' between 0 and κ, because the behavior of the real part

$$\log|\kappa - \lambda| - \log|\lambda|$$

of the numerator in passing either of the points 0 and κ is independent of the side along which the singularity is passed[5].

[5] For the same reason it is not necessary to specify in (84) on what sides these singularities are passed, although this is necessary with respect to κ in (83) where the imaginary part has not been eliminated.

Finally, if β in (85) is replaced by a large circle $|\lambda| = R$, the validity of (84) follows from the fact that (85) tends to zero if R tends to infinity because the integrand is of the order of magnitude of λ^{-2}.

The real part of the integral in (83) accordingly is

$$\Re J = \int_{-1}^{1} \frac{\log|\lambda|}{(1-\lambda^2)^{\frac{1}{2}}}\, d\lambda = 2\int_0^1 \frac{\log \lambda}{(1-\lambda^2)^{\frac{1}{2}}}\, d\lambda, \tag{86}$$

or, by the transformation $\lambda = \sin x$,

$$\begin{aligned} \Re J &= 2\int_0^{\frac{\pi}{2}} \log \sin x\, dx = 2\int_0^{\frac{\pi}{2}} \log \cos x\, dx = \int_0^{\frac{\pi}{2}} \log(\sin x \cos x)\, dx \\ &= \int_0^{\frac{\pi}{2}} \log(\tfrac{1}{2}\sin 2x)\, dx = \frac{\pi}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\Re J, \end{aligned} \tag{87}$$

so that

$$\Re J = -\pi \log 2. \tag{88}$$

In order to evaluate the imaginary part $\Im J$ of (83), it is necessary to specify on which side the singularity κ is passed by the integration variable λ. In fact, both cases need to be considered; the passage of λ "over" κ for values of κ on the first part of the path of integration γ_r' of κ in (36), where κ goes along the real axis from r to 1; and the passage of λ "under" κ for values of κ on the second part of its path γ_r', from 1 back to r. If the upper sign in the following formulae relates to the first of these two cases, we have

$$\Im J = \mp \pi i \int_\kappa^1 \frac{d\lambda}{(1-\lambda^2)^{\frac{1}{2}}} = \mp \pi i \operatorname{arc\,cos} \kappa, \tag{89}$$

and, from (88) and (89), we find for the last member in (82)

$$e^{-\frac{1}{2}TJ/\pi} = 2^{\frac{1}{2}T} e^{\pm\frac{1}{2}Ti \operatorname{arc\,cos}\kappa} \tag{90}$$

Writing

$$\operatorname{arc\,cos}\kappa = \alpha, \quad \kappa = \cos\alpha, \quad e^{\frac{1}{2}Ti\alpha} - e^{-\frac{1}{2}Ti\alpha} = 2i \sin \tfrac{1}{2}T\alpha, \tag{91}$$

we find the following approximation for $\bar{h}(\bar{r})$ by inserting (90) in (36) as indicated in (82):

$$\bar{h}(\bar{r}) \sim \frac{(\frac{1}{2}T-1)2^{\frac{1}{2}T}}{\pi} \int_0^{\operatorname{arc\,cos}\bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-2} \sin\tfrac{1}{2}T\alpha \sin\alpha\, d\alpha \tag{92}$$

Calculations of the distribution function and of its percentiles will be much simpler for this approximation than for the exact function.

In the case of $r = m/p$ in which no circular arrangement is made a slight complication arises. The characteristic values κ_l given in (74) are again ordinates of the cosine function at equidistant points, but they do not cover a complete period or half-period of this function. Probably the most accurate pro-

cedure would be to replace the limits of integration in (83) by $\cos [\frac{1}{2}\pi/(T + 1)]$ and $\cos [(T + \frac{1}{2})\pi/(T + 1)]$, so as to have each discrete integral value of t in (74) contribute an interval $(t - \frac{1}{2}, t + \frac{1}{2})$ of unit length to the range of the rectangularly distributed variable τ now defining the distribution of $\lambda = \cos [\pi\tau/(T + 1)]$; while making such an adjustment in the numerical factor in (80) that the equivalent of (81) with the new limits of integration is satisfied. However, the evaluation of (83) and the simplicity of the result essentially rest on the fact that the limits of integration coincide with singularities of the integrand. In these circumstances a rather simple result can again be obtained by introducing two further changes which very nearly compensate each other. The first change is the arbitrary extension of the limits of integration to what they are in (83), while increasing the numerical factor in (80) in such a manner that the integral in (81) will be $T + 1$ instead of T. This leaves the described contributions of the discrete values of t in (74) to the range of τ unaffected, but adds to that range the two intervals $(0, \frac{1}{2})$ and $(T + \frac{1}{2}, T)$ of half a unit length not representing anything that was already present. This can be largely offset by introducing two additional discrete values $t = 0$ and $t = T + 1$, each with the negative weight $-\frac{1}{2}$, if the weight of all other discrete values is considered to be $+1$. Instead of (82) we then have

$$e^{-\frac{1}{2}\sum_{t=1}^{T} \log (\kappa-\kappa_t)} \tag{93}$$

$$\sim \exp\left[-\frac{T+1}{2\pi}\int_{-1}^{1}\frac{\log(\kappa-\lambda)}{(1-\lambda^2)^{\frac{1}{2}}}\,d\lambda + \tfrac{1}{4}\log(\kappa-1) + \tfrac{1}{4}\log(\kappa+1)\right]$$

If this expression is inserted in (36) with γ_r constricted to γ_r', the argument of

$$e^{(\frac{1}{4})\log(\kappa-1)} = (\kappa - 1)^{1/4} \tag{94}$$

is $-\pi i/4$ when κ goes from r to 1, and $\pi i/4$ when κ returns from 1 to r. On account of

$$\begin{aligned} (1 - \kappa^2)^{1/4} &= \sin^{\frac{1}{2}} \alpha, \\ e^{\frac{1}{2}(T+1)i\alpha-\pi i/4} - e^{-\frac{1}{2}(T+1)i\alpha+\pi i/4} &= 2i \sin [\tfrac{1}{2}(T + 1)\alpha - \pi/4], \end{aligned} \tag{95}$$

the result now is

$$\begin{aligned} h(r) \sim{}& \frac{(\frac{1}{2}T - 1)2^{\frac{1}{2}T+1}}{\pi} \\ &\cdot \int_0^{\operatorname{arc}\cos r} (\cos \alpha - r)^{\frac{1}{2}T-2} \sin [\tfrac{1}{2}(T + 1)\alpha - \pi/4] \sin^{3/2} \alpha \, d\alpha. \end{aligned} \tag{96}$$

It is not necessary to prove by direct integration that the conditions equivalent to (37) are satisfied by the approximate expressions (92) and (96). This follows from the fact that the difference of 2 between the degrees in κ of the numerator and the denominator in (36) is preserved by the substitutions (82) and (93);

that the numerical value of the limit for $\kappa \to \infty$ of κ^2 times the integrand in (36) is not changed; and that no singularities outside the segment $-1 \leqq \kappa \leqq 1$ of the real axis are introduced.

There is, of course, a certain degree of distortion involved in replacing the exact distribution functions by the smooth approximations derived. Such distortion is most serious in so far as it occurs at the tails of the distribution, where the usual significance limits are located. For instance, the exact distribution of $\bar{r}$ is asymmetric if T is odd, and ranges from $\cos [(T - 1)\pi/T]$ to $+1$, whereas the smooth approximation is symmetric and ranges from -1 to $+1$. In the case of r both the exact distribution and the approximation are symmetric, but the former ranges from $\cos [T\pi/(T + 1)]$ to $\cos [\pi/(T + 1)]$, the latter from -1 to $+1$. However, this difference is to some extent compensated by a curious anomaly in the function (96) This function actually dips below zero on symmetrically placed small intervals adjoining -1 and $+1$, the length of which is of the order of the difference $1 - \cos [\pi/(T + 1)]$ between unity and the highest characteristic value. Percentiles must therefore be counted on both sides from two points absolutely smaller than unity, defined by requiring that the small parts of the area "under" the curve (95) outside these points are algebraically zero each.

These distortions have importance only for small values of T. Anderson finds ([6] p. 52) that the exact function $\bar{h}(\bar{r})$ is symmetrical within three-decimal accuracy for all values of $T \geqq 11$ (the modal value $\bar{h}(0)$ for $T = 11$ is about 1.27). There are in the case of $\bar{r}$ three characteristic values $\bar{\kappa}_t$ exceeding the 95-percentile as given by Anderson for $T = 7$; 5 for $T = 13$; 11 for $T = 25$. Corresponding numbers for the 99-percentile are 3 for $T = 13$; 9 for $T = 25$; 17 for $T = 45$ These numbers suggest that the approximations (92) and (96) will provide good significance limits long before the normal approximation is acceptable. Accurate calculations will be needed to find out from what value of T onward the approximations can safely be substituted for the exact distributions.

REFERENCES

[1] J. TINBERGEN, *Business cycles in the United States of America 1919–1932*, League of Nations, Geneva, 1939.

M. M. FLOOD, "Recursive methods in business-cycle analysis," *Econometrica*, Vol. 8 (1940) p. 333.

[2] H. WOLD, *A study in the Analysis of Stationary Time Series*, Uppsala, 1938.

[3] S. S. WILKS, *Statistical Inference*, Ann Arbor, 1937

[4] J. VON NEUMANN, "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat*, Vol. 12(1941), pp. 367–395.

[5] J. VON NEUMANN, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13(1942), pp. 86–88.

[6] R. L. ANDERSON, *Serial Correlation in the Analysis of Time Series*, unpublished thesis, Iowa State College, 1941.

A GENERALIZED ANALYSIS OF VARIANCE

BY FRANKLIN E. SATTERTHWAITE

University of Iowa and Aetna Life Insurance Company

The analysis of variance is a statistical technique whose fields of application are only beginning to be explored. A few simple standard designs appear in the literature and a great deal has been done with them. However, if the applied statistician limits himself to such standard designs, he soon finds that many of his problems are receiving inadequate or inappropriate treatment. The writer has found this particularly true in his own field where most of the raw data are in the nature of frequencies or averages which lack homogeneity of variance. Also the nature of the problem usually indicates the use of weighted averages rather than simple averages and sometimes part of the data are missing.

The purpose of this study is to examine the fundamental principles underlying analysis of variance designs and to show how designs may be constructed and applied to practically any data which can be assumed to be normally distributed.

1. Test of independence. In the analysis of variance we calculate two or more statistics of the types,

$$\chi^2 = \Sigma(x_i - m_i)^2,$$
$$\chi^2 = \Sigma\theta_i^2 .$$

The x_i's are considered to be independent variables from a normal population. The m_i's and the θ_i's are homogeneous linear functions of the x_i's. Heretofore the demonstration of the independence of the χ^2's used has only been made for certain special θ_i's and m_i's. To make our analysis general we shall let our θ_i's be general homogeneous linear functions of the x_i's and we shall define our m_i's through certain linear homogeneous restrictions.

Let us define Chi-square as

$$\chi^2 = \Sigma(x_i - m_i)$$

where the x_i's are independent normally distributed variables with mean zero and unit variance. We also define certain linear functions of the x_i's,[1]

$$\theta_j = a_{ji}x_i , \qquad j = 1, 2, \cdots s, \tag{1}$$

which we shall assume to have been orthogonalized.[2] To define the m_i's we make use of the linear restrictions

[1] A repeated lower case subscript will always indicate summation with respect to that subscript All subscripts range from 1 to n unless otherwise specified. The Kronecker Delta, δ_{ij}, equals one or zero depending on whether i equals or does not equal j.

[2] The θ_j's are orthogonal if $a_{ik}a_{jk} = \delta_{ij}$. Any algebraically independent set may be

$$a_{ji}(x_i - m_i) = 0, \qquad j = 1, 2, \cdots s, \tag{2}$$

or

$$a_{ji}m_i = a_{ji}x_i = \theta_j\,.$$

This system has an $(n - s)$-infinitude of solutions and we should not expect all of these to be suitable for our purposes. For reasons which will appear later we shall choose the single solution,

$$m_k = a_{jk}\theta_j = a_{jk}a_{ji}x_i\,, \qquad j = 1, \cdots s. \tag{3}$$

This is the solution which follows if we complete the system (2) with $n - s$ additional linear restrictions on the m_i's which are homogeneous and which form an orthogonal set with (2). Thus

$$a_{ji}m_i = \theta_j\,, \qquad j = 1, \cdots s,$$

$$a_{ji}m_i = 0, \qquad j = s + 1, \cdots n.$$

This is consistent with standard analysis of variance designs. For the usual one way analysis, we have

$$\sum_{i=1}^{r} \frac{1}{r^{1/2}} m_{ji} = \sum \frac{1}{r^{1/2}} x_{ji} \qquad j = 1, \cdots s, \tag{4}$$

which yield a solution according to (3),

$$m_{ji} = \frac{1}{r^{1/2}} \frac{1}{r^{1/2}} \Sigma_i x_{ji}\,, \qquad j = 1, \cdots s.$$

The additional homogeneous restrictions in this case might have been taken as

$$m_{j1} = m_{j2} = \cdots = m_{jr}\,, \qquad j = 1, \cdots s,$$

which are orthogonal to (4) and may be easily orthogonalized among themselves.

Substituting the values of the m_i's obtained in (3) into Chi-square, we obtain,

$$\begin{aligned}
\chi^2 &= (x_i - m_i)(x_i - m_i) \\
&= (\delta_{ik} - a_{ji}a_{jk})x_k(\delta_{il} - a_{mi}a_{ml})x_l\,, \qquad j, m = 1, \cdots s, \\
&= (\delta_{kl} - a_{mk}a_{ml} - a_{jl}a_{jk} + \delta_{jm}a_{jk}a_{ml})x_kx_l \\
&= (\delta_{kl} - a_{jk}a_{jl})x_kx_l\,.
\end{aligned}$$

---

replaced by an equivalent orthogonal set. Thus, if θ_2 is not orthogonal to θ_1, it may be replaced by $\theta_2^1 = \theta_2 + k\theta_1$, where k is determined by

$$a_{1j}(a_{2j} + ka_{1j}) = 0$$

or $k = -a_{1j}a_{2j}/a_{1j}a_{1j}$.

The condition $\Sigma a_{1j}^2 = \Sigma a_{2j}^2 = 1$ can always be met by simple division.

The sum of the squares of the θ_j's is

$$\Sigma\theta_j^2 = \theta_j\theta_j, \qquad j = 1, 2, \cdots s,$$
$$= a_{jk}a_{jl}x_kx_l.$$

Therefore we have the relation,

$$\chi^2 + \Sigma\theta_j^2 = \delta_{kl}x_kx_l = \Sigma x_i^2 \tag{5}$$

The rank, R_j, of each θ_j^2 is obviously equal to unity since it is the square of a linear form. The rank, R_0, of χ^2 is at least equal to $n - s$ since the rank of the right hand side of (5) is n. Also, R_0 can not be greater than $n - s$ since,

$$a_{jk}(\delta_{kl} - a_{ik}a_{il}) = a_{jl} - \delta_{ji}a_{il}, \qquad i, j = 1, \cdots s,$$
$$= 0$$

gives s independent relations between the rows of its coefficient matrix. Therefore we have the relation,

$$R_0 + R_1 + \cdots + R_s = n. \tag{6}$$

The two conditions, (5) and (6), are sufficient[3] conditions for χ^2 and the θ_j^2's each to be independent of the others and each to be distributed as is Chi-square with the number of degrees of freedom equal to its rank.

2. Adjustment of data. The above development is not general enough for many practical problems. We do not always have given data, y_i, which are normally distributed about a mean zero with unit (or homogeneous) variance. Of course if the means, $\bar{m}_i$, and variances, σ_i^2, are known, we may make the transformation,

$$x_i = \frac{y_i - \bar{m}_i}{\sigma_i} \tag{7}$$

and apply our theory in a straight forward manner. We shall now check the effect on our analysis if the $\bar{m}_i$'s and σ_i's are determined, in part at least, from our data, the y_i's.

Let us assume that the x_i's of (7) are normally and independently distributed variables about a mean zero and with unit variance. Let us also define certain linear orthogonal functions of the first r of the θ_j's by

$$\begin{aligned} \phi_k &= b_{kj}\theta_j = b_{kj}a_{ji}x_i & k &= 1, 2, \cdots q, \\ &= b_{kj}a_{ji}\left(\frac{y_i - \bar{m}_i}{\sigma_i}\right) & j &= 1, 2, \cdots r. \end{aligned} \tag{8}$$

We next form the characteristic function of the joint distribution of χ^2, of $\Sigma_1^r\theta_j^2$, of $\theta_{r+1}^2, \cdots \theta_s^2$, and of $\phi_1, \cdots \phi_q$. This is

[3] See A. T. Craig, "On the independence of certain estimates of variance," *Annals of Math Stat.*, Vol. 9(1938), pp. 46–55.

$$\Phi(t, u, v_{r+1}, \cdots, v_s, w_1, \cdots w_q)$$

$$= K \int \cdots \int \exp [it\chi^2 + iu\Sigma_1^r \theta_j^2 + i\Sigma_{r+1}^s v_j \theta_j^2 + i\Sigma_1^q w_j \phi_j - \tfrac{1}{2}\Sigma_1^n x_j^2]\, dx_n \cdots dx_1 .$$

The conditions (5) and (6) are sufficient[4] for there to exist an orthogonal transformation of the x_i's which will convert

$$\begin{aligned} &\theta_j \quad \text{to} \quad \theta_j , && j = 1, \cdots s, \\ &\chi^2 \quad \text{to} \quad \Sigma_{s+1}^n \theta_j^2 , \\ &\Sigma_1^n x_j^2 \quad \text{to} \quad \Sigma_1^n \theta_j^2 , \\ &dV = \Pi\, dx_j \quad \text{to} \quad \Pi\, d\theta_j . \end{aligned}$$

The characteristic function then takes the form,

$$\Phi = K \left\{ \Pi_1^r \int \exp \left[-\tfrac{1}{2}(1 - 2iu) \left(\theta_j - \frac{iw_k b_{k,j}}{1 - 2iu} \right)^2 \right] d\theta_j \right\}$$

$$\{ \Pi_1^q \exp [w_k^2/2(1 - 2iu)] \}$$

$$\left\{ \Pi_{r+1}^s \int \exp [-\tfrac{1}{2}(1 - 2iv)\theta_j^2]\, d\theta_j \right\}$$

$$\left\{ \Pi_{s+1}^n \int \exp [-\tfrac{1}{2}(1 - 2it)\theta_j^2]\, d\theta_j \right\},$$

where

$$\Sigma(w_k b_{kj})^2 = \Sigma w_k^2 ,$$

since the b_{kj}'s are orthogonal.

At the beginning of this section we stated that we wished in some way to use our data, the y_i's, to estimate the $\dot{m}_i$'s and the σ_i's. A suitable method is to restrict the ϕ functions, (8), to zero.

Our problem thus reduces to finding the distribution of the "array" in our joint distribution for which

$$\phi_1 = \phi_2 = \cdots = \phi_q = 0.$$

Except for perhaps a constant factor, the characteristic function of the distribution of such an array is obtained from Φ by integrating out the w_k's.[5] Thus, on performing the integrations, we have,

---

[4] See A. T. Craig, *ibid.*

[5] This is easily seen since if one passes from the characteristic function to the joint distribution, equates the ϕ_k's to zero, and then passes back to the characteristic function, all the integrations except the above appear in pairs of the form

$$\frac{1}{2\pi} \int e^{itx} \int e^{-itx} \Phi\, dt\, dx,$$

which leave Φ unchanged.

$$\Phi'(t, u, v_{r+1}, \cdots, v_s) = K\{(1 - 2iu)^{-(r-q)/2}\}\{\Pi_{r+1}^{s}(1 - 2iv_j)^{-1/2}\}$$
$$\cdot\{(1 - 2it)^{-(n-s)/2}\}.$$
$$= \Phi_0(u)\{\Pi_{r+1}^{s}\Phi_j(v_j)\}\Phi_n(t),$$

which shows that $\Sigma_1^r \theta_j^2$, $\theta_{r+1}^2, \cdots \theta_s^2$, and χ^2 are each independent of the others and that each is distributed according to the Chi-square distribution with $r - q$, 1, $\cdots$ 1, and $n - s$ degrees of freedom respectively.

3. **Numerical application.** The developments of the preceding sections have been abbreviated to cover technical points alone. We shall now take a definite practical problem and see how we may work out its solution with the aid of the above techniques.

In Table I are given the losses, the exposures (in car years) and the indicated pure premiums from the Massachusetts Statutory Liability automobile insurance experience for four towns and for three different classes of cars. (To illustrate the effect of missing items, the data for town D, class W, and for town C, class Y, have been omitted.) Our problem is to determine if there is a significant variation in the indicated pure premium between the different towns and between the different classes of cars.

Our first problem is to set up a normally distributed variable about a mean zero and with homogeneous variance. The true mean, $\bar{m}_i$, of the distribution of the indicated pure premiums, P_i, is unknown. Under the hypothesis that the different towns and classes of cars are homogeneous with each other, we may assume that the $\bar{m}_i$'s are all equal. We may estimate their value by using the combined indicated P for the whole territory, which is \$32.44. By a preliminary argument, which need not concern us here, we show that the variance, σ_i^2, of an indicated pure premium is inversely proportional to the exposure, E_i, on which it is based but the constant of proportionality is unknown. If we now make the assumption that the indicated pure premiums are normally distributed, we may convert them to the form

$$x_i = \frac{P_i - 32.44}{1/E_i^{1/2}}$$

which will be normally distributed about a mean zero with homogeneous variance. We have calculated these statistics and entered them in the table. Because the expected value of P_i, \$32.44, was estimated from our data, the x_i's are subject to the single homogeneous linear restriction,

$$0 = \Sigma(L_i - \bar{P}E_i)$$
$$= \Sigma E_i^{1/2}\frac{P_i - \bar{P}}{1/E_i^{1/2}} \tag{9}$$
$$= \Sigma E_i^{1/2}x_i.$$

TABLE I

i	Class	Town	Exposure E_i	Losses L_i	Premium P_i	$\sqrt{E_i}$	$\sqrt{E_i}(P_i - \bar{P}) = x_i$	Restrictions, (j) 1	2	3	4	5′	6′	5	6	7
1	W	A	4,800	154,500	32.19	69.2	−17.3	69.2				37 1	3 4	69.2		
2		B	3,041	102,900	33 85	55 1	77 6		55 1			30 5	1 6	55 1		
3		C	4,012	147,500	36 77	63.4	274 5			63 4		32.2	−4 9	63 4		
$\bar{5}$		T	11,853	404,900	34.16	108.9	(187 2)									
4	X	A	4,188	131,800	31 46	64.7	−63.4	64.7				−30 0	13 2		64.7	
5		B	2,995	83,900	28.02	54 8	−242 2		54.8			−24.4	10.1		54 8	
6		C	4,141	131,700	31.79	64.4	−41 9			64.4		−31.7	4 9		64 4	
7		D	1,004	29,700	29 60	31.6	−89.7				31.6		9.2		31 6	
$\bar{6}$		T	12,328	377,100	30 59	111.0	(−205.4)									
8	Y	A	1,365	57,500	42.14	37.0	358 9	37 0				−17 1	−29.4			37.0
9		B	778	16,300	20 93	27 9	−321.1		27.9			−12.4	−22.8			27 9
10		D	406	11,200	27.67	20.2	−96.4				20 2		−14 3			20.2
7		T	2,549	85,000	33 35	50.5	(46.0)									
$\bar{1}$	T	A	10,353	343,800	33 21	101 8	(77.6)									
$\bar{2}$		B	6,814	203,100	29.81	82.5	(−217.4)									
$\bar{3}$		C	8,153	279,200	34 25	90 3	(163 0)									
$\bar{4}$		D	1,410	40,900	29 01	37.5	(−128.9)									
Total			26,730	867,000	32.44		(0)									
1′	Σa_{ij}^2						395,360	10,340	6,810	8,170	1,410	6,290	2,010			
2′	$k_{5j} = \Sigma a_{5i} a_{ij} / \Sigma a_{ij}^2$							.464	.446	492	0					
3′	$k_{6j} = \Sigma a_{i6} a_{ij} / \Sigma a_{ij}^2$							404	.440	.507	.708	− 846				
4′	Σa_{ij}							170 9	137.8	127 8	51.8	−15 8	−29 0	187 7	215 5	85 1
5′	$\theta_j^2 = x_i^2$ or $= (\Sigma a_{ij} x_i)^2 / \Sigma a_{ij}^2$						(173,051)	6,022	47,263	26,568	16,615	49,006	27,576	(35,044)	(42,189)	(2,116)

The next step is to express the indicated pure premiums for each town and for each class of car as θ_j's as defined in equation (1). For town A we have an indicated pure premium of \$33.21 when all classes of cars are combined. This breaks down as follows:

$$
\begin{aligned}
33.21 &= \Sigma E_i P_i/\Sigma E_i, \qquad i = 1, 4, 8, \\
&= [\Sigma E_i(x_i/E_i^{1/2} + 32.44)]/\Sigma E_i \\
&= \Sigma(E_i^{1/2}/\Sigma E_i)x_i + 32.44.
\end{aligned}
$$

Dividing this by the square root of the sum of the squares of the coefficients, we obtain,

$$
\begin{aligned}
\theta_1 &= (\Sigma E_i)^{1/2}(33.21 - 32.44), \qquad i = 1, 4, 8, \\
&= \Sigma(E_i^{1/2}/(\Sigma E_i)^{1/2})x_i ,
\end{aligned}
\tag{10}
$$

which is of the form of (1). We have entered the coefficients of θ_1 (except for the common denominator, $(\Sigma E_i)^{1/2}$, whose square is entered on line (1')) under Restriction (1) in the table. Similarly, we have entered the values for the other towns under Restrictions (2), (3), and (4). The values for the classes of cars are entered under (5), (6), and (7).

The next step is to orthogonalize the θ_j's. The first four have no common elements so they are orthogonal by inspection. To make θ_5 orthogonal to θ_1 we must add to θ_5,

$$k_{51} = -\Sigma a_{i5}a_{i1}/\Sigma a_{i1}^2$$

times θ_1. This and similar coefficients for making θ_5 orthogonal to θ_2, θ_3, and θ_4 are entered on line (2'). We may now replace θ_5 by the equivalent $\theta_{5'}$ by the formula

$$a_{i5'} = a_{i5} + k_{51}a_{i1} + k_{52}a_{i2} + k_{53}a_{i3} + k_{54}a_{i4} . \tag{11}$$

Similar k's for θ_6 are entered on line (3') and θ_6 is replaced by $\theta_{6'}$. θ_7 should be ignored since it is algebraically dependent on the other θ_j's:

$$\theta_7 = \theta_1 + \theta_2 + \theta_3 + \theta_4 - \theta_5 - \theta_6 .$$

Note that on line (4') we have entered Σa_{ij} for checking the calculation (11).

We next calculate the θ_j^2's according to the formula,

$$\theta_j^2 = [\Sigma a_{ij}x_i]^2/\Sigma a_{ij}^2 .$$

Note that for this particular design all the θ_j's except $\theta_{5'}$ and $\theta_{6'}$ are numerically equal to the corresponding x_j's (enclosed in parentheses).

Returning to equation (9), we see that it is equivalent to either of the following restrictions on the θ_j's:

$$E_1^{1/2}\theta_1 + E_2^{1/2}\theta_2 + E_3^{1/2}\theta_3 + E_4^{1/2}\theta_4 = 0$$

or

$$E_{\bar{5}}^{1/2}\theta_5 + E_{\bar{6}}^{1/2}\theta_6 + E_{\bar{7}}^{1/2}\theta_7 = 0.$$

Therefore we may conclude that

$$S_1^2/\sigma_x^2 = (\theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2)/\sigma_x^2 = 96{,}469/\sigma_x^2$$

is distributed as is Chi-square with three degrees of freedom. Also we may conclude that

$$S_2^2/\sigma_x^2 = (\theta_5^2 + \theta_6^2 + \theta_7^2)/\sigma_x^2 = 79{,}349/\sigma_x^2\,,$$

is distributed as is Chi-square with two degrees of freedom. Note that we have not proved, and indeed it is not so, that S_1^2 and S_2^2 are independent.

We have yet to obtain our interaction sum of squares Equation (5) is of assistance, here giving,

$$S_3^2/\sigma_x^2 = [\Sigma x_i^2 - (\theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2 + \theta_{5'}^2 + \theta_{6'}^2)]/\sigma_x^2$$

$$= \frac{395{,}360 - 173{,}051}{\sigma_x^2} = \frac{222{,}309}{\sigma_x^2}\,.$$

This is distributed as is Chi-square with $10 - 6 = 4$ degrees of freedom. Also it is independent of S_1^2 and of S_2^2.

Lastly we form the variance ratios

$$F_1 = \frac{96.469/3}{222{,}309/4} = 0.58,$$

$$F_2 = \frac{79{,}349/2}{222{,}309/4} = 0.71,$$

which are not significant.

We therefore conclude that as far as the present data and analysis show, we have no reason to believe that these three classes of cars and these four towns are not all subject to the same true premium rate.

DISTRIBUTIONS IN STRATIFIED SAMPLING

By Paul H. Anderson

University of Illinois

1. **Introduction.** In this paper, distributions of means and standard deviations will be derived for random and stratified samples. It is not necessary to define random sampling here, for one may find it defined in any elementary text. If before drawing the sample from a population π, it is divided into several strata $\pi_1, \pi_2, \cdots, \pi_s$, and the sample Σ is composed of s partial samples $\Sigma_1, \Sigma_2, \cdots, \Sigma_s$ each drawn with or without replacement from the strata; and if the sizes m_i of the partial samples are proportionate to the sizes M_i of corresponding strata, i e., $m_i = kM_i$, then the sample which is obtained in this manner is a stratified sample. When the sizes of the partial samples are not proportionate to the sizes of the corresponding strata, the distributions of means and standard deviations will differ from the distributions obtained when the sizes of the partial samples are proportionate to the sizes of the corresponding strata. This will be shown in the sections that follow.

The distributions of means and standard deviations from well-known populations for stratified and random samples will be derived and compared, as to scatter and symmetry. It should be remembered even though stratification has little to recommend its use, in some cases, over random sampling, the impossibility of obtaining random samples makes its use necessary. Since most of the problems with which the practical statistician is confronted are of the kind which make random sampling difficult or even impossible, stratified sampling is being investigated by many research workers.

2. **The distribution of means and standard deviations for samples of two drawn from any population having a continuous frequency function.** Let $f(x)$ be a continuous frequency function whose mean is zero, and for $a \leq x \leq b$, let $f(x) > 0$, elsewhere let $f(x) = 0$. We select a sample of two elements (x_1, x_2) which can be represented by a point in a square of side $b - a$, as point P in Fig. 1. It is well known that the probability of getting a sample point in the element of area $dx_1\,dx_2$ is $f(x_1)f(x_2)\,dx_1\,dx_2$. The probability of getting a value of $\bar{x}$ (mean) less than the value of the mean represented by a point on the line RT (Fig. 1) whose equation is $x_1 + x_2 = 2\bar{x}$, is given by

$$(1)\qquad \int_a^{2\bar{x}-a} dx_1 \int_a^{2\bar{x}-x_1} dx_2 f(x_1)f(x_2).$$

The distribution of $\bar{x} \leq \frac{1}{2}(a + b)$ is

$$(2)\qquad 2\int_a^{2\bar{x}-a} f(x_1)f(2\bar{x} - x_1)\,dx_1,$$

which is obtained by differentiating (1) with respect to $\bar{x}$. For all values $\bar{x} \geq \frac{1}{2}(a + b)$, we must use another equation which we shall now derive similarly. The probability of obtaining a mean less than the mean of any point on $R'T'$ (Fig. 1) is

$$1 - \int_{2\bar{x}-b}^{b} dx_1 \int_{2\bar{x}-x_1}^{b} dx_2 f(x_1) f(x_2).$$

Differentiating this expression, we obtain

$$(3) \qquad 2 \int_{2\bar{x}-b}^{b} f(x_1) f(2\bar{x} - x_1)\, dx_1 .$$

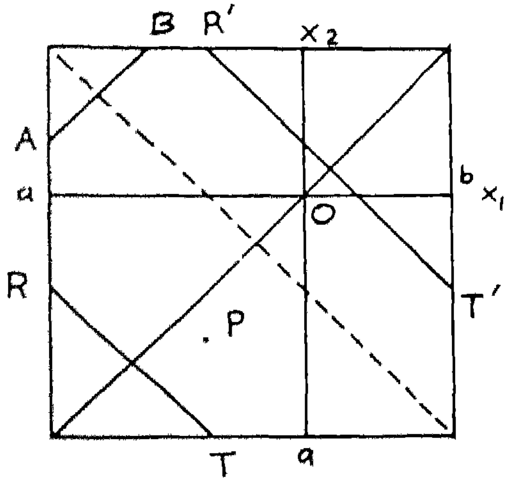


FIG. 1

The distribution of means is given by (2) and (3). Let us apply the theorem to the rectangular population

$$f(x) \begin{cases} = 1, & -\frac{1}{2} \leq x \leq \frac{1}{2}, \quad a = -\frac{1}{2}, \quad b = \frac{1}{2}, \\ = 0 & \text{elsewhere.} \end{cases}$$

Substituting in (2) and (3) respectively, the results obtained are

$$g(\bar{x}) \begin{cases} = 2(1 + 2\bar{x}), & \text{for } \bar{x} \leq 0, \\ = 2(1 - 2\bar{x}), & \text{for } \bar{x} \geq 0. \end{cases}$$

J. O. Irwin [1] and Philip Hall [2] obtained these results also but by different methods. However, the distribution of $2\bar{x}$ was known to Laplace and other earlier writers.

From Fig. 1, it is seen that the probability of obtaining a value of S (standard deviation), less than the value of S on AB whose equation is $x_2 - x_1 = 2S$ is

$$1 - 2\int_a^{b-2S} dx_1 \int_{2S+x_1}^{b} dx_2 f(x_1) f(x_2).$$

Upon differentiating this expression with respect to S, we obtain

(4) $$h(S) = 4\int_a^{b-2S} f(x) f(2S + x)\, dx.$$

For the rectangular population $h(S) = 4(1 - 2S)$. This result agrees with that found by P. R. Rider [3].

3. Sampling from a rectangular population. Let the rectangular population be $f(x) = 1$, for $0 \leq x \leq 1$, elsewhere $f(x) = 0$. From this population we

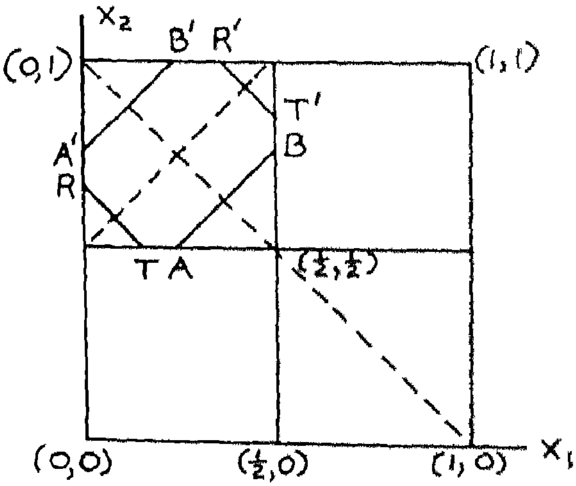


Fig. 2

select a stratified sample of two elements which is chosen so that $0 \leq x_1 \leq \frac{1}{2}$ and $\frac{1}{2} \leq x_2 \leq 1$. The probability of obtaining a mean less than the mean of any point on the line $R'T'$ (Fig. 2) whose equation is $x_1 + x_2 = 2\bar{x}$, is

$$4\int_0^{2\bar{x}-\frac{1}{2}} dx_1 \int_0^{2\bar{x}-x_1} dx_2 = 4(2\bar{x}^2 - \bar{x} + \tfrac{1}{8}).$$

Similarly, the probability of obtaining a mean less than the mean of any point on $R'T'$ (Fig. 2) whose equation is $x_1 + x_2 = 2\bar{x}$, is

$$1 - 4\int_{2\bar{x}-1}^{\frac{1}{2}} dx_1 \int_{2\bar{x}-x_1}^{1} dx_2 = 12\bar{x} - 8\bar{x}^2 - \tfrac{7}{2}.$$

Differentiating the right-hand side of the above two equations with respect to $\bar{x}$, we get the distribution of means of stratified samples of two elements from a rectangular distribution function to be

$$(5)\qquad g(x) = \begin{cases} 16x - 4, & \text{for } \frac{1}{4} \le x \le \frac{1}{2}, \\ 12 - 16x, & \text{for } \frac{1}{2} \le x \le \frac{3}{4}. \end{cases}$$

The distribution of means for random samples of two elements from the same rectangular population is

$$(6)\qquad g(x) = \begin{cases} 4x, & \text{for } 0 \le x \le \frac{1}{2}, \\ 4 - 4x, & \text{for } \frac{1}{2} \le x \le 1. \end{cases}$$

Upon examining (5) and (6) we see that

A. The stratified sample means are more stable than the random means.

B. The random sample means and the stratified sample means are both distributed symmetrically.

C. The range of the random means is twice the range of the stratified means.

It remains now to find the distributions of the standard deviations for samples of two elements where one element is selected from each half of the population. All points on AB (Fig. 2) have the same standard deviation. Furthermore the equation of AB is $x_2 - x_1 = 2S$. The probability of obtaining a standard deviation less than the standard deviation of any point on AB (Fig. 2) is

$$4\int_{\frac{1}{2}}^{\frac{1}{2}+2S} dx_2 \int_{x_2-2S}^{\frac{1}{2}} dx_1 = 8S^2.$$

Furthermore, the probability of getting a standard deviation less than the standard deviation on the line $A'B'$ (Fig. 2) of which the equation is $x_2 - x_1 = 2S$, is

$$1 - 4\int_0^{1-2S} dx_1 \int_{2S+x_1}^{1} dx_2 = -1 - 8S^2 + 8S.$$

Differentiation of the right-hand side of the above two equations with respect to S yields the distribution of standard deviations of stratified samples of two elements from a rectangular distribution function to be

$$(7)\qquad h(S) = \begin{cases} 16S, & \text{for } 0 \le S \le \frac{1}{4}, \\ 8 - 16S, & \text{for } \frac{1}{4} \le S \le \frac{1}{2}. \end{cases}$$

The distribution of the standard deviations for random samples of two elements is

$$(8)\qquad h(S) = 4(1 - 2S), \qquad \text{for } 0 \le S \le \tfrac{1}{2}.$$

From (7) and (8) it is easily seen that:

A. The range of the standard deviations for stratified and random samples is the same.

B. The distribution of standard deviations for random samples of two elements is skewed, but the distribution of the standard deviations for stratified samples of two elements is symmetrical.

If we take a random sample of two elements from the rectangular population on the interval $-\frac{1}{2} \leq x \leq \frac{1}{2}$, then Student's ratio $t = \sqrt{2}\bar{x}/\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2}$ will have the distribution

$$F(t) \begin{array}{ll} = 1/2(t-1)^2 & \text{for } t \leq 0, \\ = 1/2(t+1)^2 & \text{for } t \geq 0. \end{array}$$

This result was obtained by Laderman [7] and others. According to the reasoning used by Laderman, the probability of getting a value of t less than the value on OS is (for stratified samples of two elements)

$$4 \int_{-\frac{1}{2}}^{0} dx_1 \int_{0}^{x_1(t+1)/(t-1)} dx_2 = -\tfrac{1}{2}(t+1)/(t-1).$$

When $t \geq 0$, the probability of obtaining a value of t greater than the value on OS is

$$4 \int_0^{\frac{1}{2}} dx_2 \int_{x_2(t-1)/(t+1)}^{0} dx_1 .$$

It follows easily that the probability of getting a value of t less than the value represented on OS is for stratified samples equal to

$$1 - 4 \int_0^{\frac{1}{2}} dx_2 \int_{x_2(t-1)/(t+1)}^{0} dx_1 = 1 + \tfrac{1}{2}(t-1)/(t+1).$$

Differentiating the right-hand side of the first and third above equations with respect to t, we find the distribution of Student's ratio for stratified samples of two elements from a rectangular population to be

$$F(t) \begin{array}{ll} = 1/(t-1)^2, & \text{for } -1 \leq t \leq 0, \\ = 1/(t+1)^2, & \text{for } \quad 0 \leq t \leq 1. \end{array}$$

Comparing the random sample and stratified sample distributions of t, we find that

A. The stratified t's are more stable than the random t's.

B. Both distributions are symmetrical.

C. The range for the stratified t's is $-1 \leq t \leq +1$, while the range for the t's obtained from random samples is $-\infty \leq t \leq +\infty$.

By means of a different method, distributions of means of stratified samples will be obtained. Let (A) and (B) be rectangular populations $f(x)$, $f(y)$ respectively, with positive values on the interval 0, 1. From the rectangular population (A) select a stratified sample of two elements x_1 and x_2 such that $0 \leq x_1 \leq \frac{1}{2}$, $\frac{1}{2} \leq x_2 \leq 1$. Then the probability of getting a sample point in the element of area $dx_1\,dx_2$ is $4\,dx_1\,dx_2$. Now let $y_1 = 2x_1$ (change of unit of measurement), $y_2 = 2x_2 - 1$ (change of unit of measurement and translation). Then $4\,dx_1\,dx_2 = dy_1\,dy_2$. We have also that $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$. With re-

spect to the distribution of the means, a stratified sample of two elements from (A) is the same as a random sample of two elements from (B). Now the means for random samples of two elements from (B) have the distribution $g(\bar{y})$ which is really expression [illegible] with $\bar{y}$ substituted for $\bar{x}$. Furthermore, we have $\bar{y} = \frac{1}{2}(y_1 + y_2) = \frac{1}{2}(2x_1 + 2x_2 - 1) = 2\bar{x} - \frac{1}{2}$. Hence it follows readily that $g(\bar{x}) = 16\bar{x} - 4$ for $\frac{1}{4} \leq \bar{x} \leq \frac{1}{2}$, $g(\bar{x}) = 12 - 16\bar{x}$ for $\frac{1}{2} \leq \bar{x} \leq \frac{3}{4}$.

From the rectangular population (A), take a stratified sample of three elements $0 \leq x_1 \leq \frac{1}{3}$, $\frac{1}{3} \leq x_2 \leq \frac{2}{3}$, $\frac{2}{3} \leq x_3 \leq 1$. The sample points will all lie in a cube within the unit cube. Then the probability of getting a sample point in the element of volume $dx_1\,dx_2\,dx_3$ is $27\,dx_1\,dx_2\,dx_3$. Now let $y_1 = 3x_1$, $y_2 = 3x_2 - 1$, $y_3 = 3x_3 - 2$. Therefore $0 \leq y_i \leq 1$, for $i = 1, 2, 3$. Furthermore, $dy_1\,dy_2\,dy_3 = 27\,dx_1\,dx_2\,dx_3$. With respect to the distribution of the means, a stratified sample of three elements from (A) is the same as a random sample of three elements from (B). Now the means for random sample of three elements from (B) have the distribution

$$g(\bar{y}) = \begin{cases} 27\bar{y}^2/2, & \text{for } 0 \leq \bar{y} \leq \frac{1}{3}, \\ 9(6\bar{y} - 6\bar{y}^2 - 1)/2, & \text{for } \frac{1}{3} \leq \bar{y} \leq \frac{2}{3}, \\ 27(1 - \bar{y})^2/2, & \text{for } \frac{2}{3} \leq \bar{y} \leq 1. \end{cases}$$

We have also $\bar{y} = 3\bar{x} - 1$. For $\bar{y} = 0, \frac{1}{3}, \frac{2}{3}, 1$, $\bar{x} = \frac{1}{3}, \frac{4}{9}, \frac{5}{9}, \frac{2}{3}$, respectively. Hence

$$g(\bar{x}) = \begin{cases} 81(3\bar{x} - 1)^2/2, & \text{for } \frac{1}{3} \leq \bar{x} \leq \frac{4}{9}, \\ 27(54\bar{x} - 54\bar{x}^2 - 13)/2, & \text{for } \frac{4}{9} \leq \bar{x} \leq \frac{5}{9}, \\ 81(2 - 3\bar{x})^2/2, & \text{for } \frac{5}{9} \leq \bar{x} \leq \frac{2}{3}. \end{cases}$$

Thus we have found the distribution of the means for stratified samples of three elements when one element is selected from each third of the population.

From the rectangular population (A), take a stratified sample of four elements $0 \leq x_1 \leq \frac{1}{4}$, $\frac{1}{4} \leq x_2 \leq \frac{1}{2}$, $\frac{1}{2} \leq x_3 \leq \frac{3}{4}$, $\frac{3}{4} \leq x_4 \leq 1$. Again, a stratified sample of four elements from (A) (with respect to the distribution of means) is the same as a random sample of four elements from (B). The means for random samples of four elements from (B) have the distribution:

$$\text{(C)} \quad g(\bar{y}) = \begin{cases} 128\bar{y}^3/3, & \text{for } 0 \leq \bar{y} \leq \frac{1}{4}, \\ 8[1 - 24(\bar{y} - \frac{1}{2})^2 - 48(\bar{y} - \frac{1}{2})^3]/3, & \text{for } \frac{1}{4} \leq \bar{y} \leq \frac{1}{2}, \\ 8[1 - 24(\bar{y} - \frac{1}{2})^2 + 48(\bar{y} - \frac{1}{2})^3]/3, & \text{for } \frac{1}{2} \leq \bar{y} \leq \frac{3}{4}, \\ 128(1 - \bar{y})^3/3, & \text{for } \frac{3}{4} \leq \bar{y} \leq 1. \end{cases}$$

Since $\bar{y} = 4\bar{x} - \frac{3}{2}$, we have for $\bar{y} = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ respectively, $\bar{x} = \frac{3}{8}, \frac{7}{16}, \frac{1}{2}, \frac{9}{16}, \frac{10}{16}$. Hence

$$g(\bar{x}) = \begin{cases} 512(4\bar{x} - \frac{3}{2})^3/3, & \text{for } \frac{3}{8} \leq \bar{x} \leq \frac{7}{16}, \\ 32[1 - 24(4\bar{x} - 2)^2 - 48(4\bar{x} - 2)^3]/3, & \text{for } \frac{7}{16} \leq \bar{x} \leq \frac{1}{2}, \\ 32[1 - 24(4\bar{x} - 2)^2 + 48(4\bar{x} - 2)^3]/3, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{9}{16}, \\ 512(1 - 4\bar{x} + \frac{3}{2})^3/3, & \text{for } \frac{9}{16} \leq \bar{x} \leq \frac{10}{16}. \end{cases}$$

This is the distribution of the means for stratified samples of four elements (one element from each quartile). We can extend this to stratified samples of size n where one element is selected from each stratum and there are n strata. As n increases, we note that

A. The range of the means decreases.

B. The scatter of the means decreases.

C. The number of arcs in the distribution of the means increases.

Take the stratified sample of four elements (two elements from each half), $0 \leq x_1 \leq \frac{1}{2}, 0 \leq x_2 \leq \frac{1}{2}, \frac{1}{2} \leq x_3 \leq 1, \frac{1}{2} \leq x_4 \leq 1$. With respect to the distribution of the means, a stratified sample of four elements from (A) is the same as a random sample of four elements from (B). Now the means for random samples of four elements from (B) have the distribution (C). Furthermore $\bar{y} = 2\bar{x} - \frac{1}{2}$, $d\bar{y} = 2\,d\bar{x}$, and for $\bar{y} = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$, $\bar{x} = \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}$. Thus

$$g(\bar{x}) = \begin{cases} 256(2\bar{x} - \frac{1}{2})^3/3, & \text{for } \frac{1}{4} \leq \bar{x} \leq \frac{3}{8}, \\ 16[1 - 24(2\bar{x} - 1)^2 - 48(2\bar{x} - 1)^3]/3, & \text{for } \frac{3}{8} \leq \bar{x} \leq \frac{1}{2}, \\ 16[1 - 24(2\bar{x} - 1)^2 + 48(2\bar{x} - 1)^3]/3, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{5}{8}, \\ 256(\frac{3}{2} - 2\bar{x})^3/3, & \text{for } \frac{5}{8} \leq \bar{x} \leq \frac{3}{4}. \end{cases}$$

Hence the distribution of the means for stratified sample of four elements (two elements from each half) has been found.

If we take a stratified sample of six elements (three elements from each half) we find that the graph of the distribution function of the means will consist of six arcs; the range will be $\frac{1}{4} \leq \bar{x} \leq \frac{3}{4}$. Thus we see that as we take more elements from each half, the distribution becomes smoother. The number of arcs in the distribution of the means also increases. The range of the means remains the same but scatter decreases as we take more elements from each half of the population (A).

The results so far obtained are true for the rectangular population which is symmetric. In order to make further comparisons in the distributions of means and standard deviations for stratified and random samples, let us now consider a skewed distribution.

4 **Sampling from a skewed population.** Let us consider the population $f(x) = 2x, 0 \leq x \leq 1, f(x) = 0$ elsewhere. If we take random samples of two elements from this population, the points represented by each sample will lie inside the unit square. For random samples of two elements from this population the distribution of means will consist of two cubics $g(\bar{x}) = 32\bar{x}^3/3$, for $0 \leq \bar{x} \leq \frac{1}{2}$, $g(\bar{x}) = 16(3\bar{x} - 2\bar{x}^3 - 1)/3$, for $\frac{1}{2} \leq \bar{x} \leq 1$. Furthermore, the distribution of the standard deviation for random sample of two elements is a cubic: $h(S) = 16(4S^3 - 3S + 1)/3$, for $0 \leq S \leq \frac{1}{2}$. Now we consider the distribution of means for stratified samples of two elements when one element is selected from the range ($0 \leq x_1 \leq \frac{1}{2}$) which comprises one fourth of the total

population. The other element is selected from the range ($\frac{1}{2} \leq x_2 \leq 1$) which constitutes three quarters of the total population. By use of the geometric method the distribution of the stratified means is found to be

$$g(\bar{x}) \begin{cases} = 16(32\bar{x}^3 - 6\bar{x} + 1)/9, & \text{for } \frac{1}{4} \leq \bar{x} \leq \frac{1}{2}, \\ = 16(30\bar{x} - 9 - 32\bar{x}^3)/9, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{3}{4}. \end{cases}$$

The range of the stratified means is less, and the distribution is more nearly symmetrical than it is for the random means as may be seen by comparing the graphs of the two distribution functions. Thus we see that stratification gives the means greater stability. The distribution of the standard deviations of the stratified samples of two elements is:

$$h(S) \begin{cases} = 64(3S - 8S^3)/9, & \text{for } 0 \leq S \leq \frac{1}{4}, \\ = 128(4S^3 - 3S + 1)/9, & \text{for } \frac{1}{4} \leq S \leq \frac{1}{2}. \end{cases}$$

Upon comparing the distributions of the standard deviations for random and stratified samples, we observe that the random case yields a single cubic whereas the stratified case yields two cubics. The distribution obtained for the stratified case is more symmetrical than it is for the random case as may be seen by sketching the graphs of the two distribution functions. The range for both distributions is the same.

5. **Sampling from a normal population.** We shall consider a normal population F having the frequency function $e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(-\infty \leq x \leq \infty)$; and the ith moment about the mean will be μ_i. Divide this population into two equal parts F_1 and F_2 such that the frequency function of F_1 is $2e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(-\infty \leq x \leq 0)$, and the frequency function of F_2 is $2e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(0 \leq x \leq \infty)$. The ith moment of F_1 about the origin will be m_{i1}, while the ith moment about its mean will be μ_{i1}; the corresponding ith moments for F_2 will be m_{i2} and μ_{i2} respectively. In what follows M_i' will be the ith moment about the origin of the distribution sought, while M_i will be the ith moment about the mean. Furthermore, the constants β_1, β_2, κ, S_k (measure of skewness) which will be used here are defined in Elderton [8]. Finally, $E[f(x)]$ will be the expected value of $f(x)$.

If we take a random sample of n elements $x_1, x_2, \cdots, x_n$ from F_1 and a random sample of n elements $x_{n+1}, \cdots, x_{2n}$ from F_2, the $2n$ elements $x_1, \cdots, x_n, x_{n+1}, \cdots, x_{2n}$ will be a stratified sample from the population F. Let $\bar{x}_1 = (1/n) \sum_{i=1}^{n} x_i$, $\bar{x}_2 = (1/n) \sum_{i=n+1}^{2n} x_i$, and $\bar{x} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$; then $\bar{x}$ will be the mean of the stratified sample. By using Tchouproff's [6] formulae and expected values, we obtain the following values:

$$M' = E(\bar{x}) = \tfrac{1}{2}(m_{11} + m_{12}) = 0,$$

$$M_2 = E(\bar{x}^2) = (\mu_{21} + \mu_{22})/4n = (1 - m_{12}^2)/2n,$$

$$M_3 = E(\bar{x}^3) = (\mu_{31} + \mu_{32})/8n^2 = 0,$$

$$M_4 = E(\bar{x}^4) = [\mu_{41} + \mu_{42} + 3n(\mu_{21} + \mu_{22})^2 - 3(\mu_{21}^2 + \mu_{22}^2)]/16n^3,$$

$$\beta_1 = M_3^2/M_2^3 = 0, \qquad \beta_2 = M_4/M_2^2 = 3 + 4(\pi - 3)/n(\pi - 2)^2.$$

From these constants, we see that the variance of the stratified means is $(1 - 2/\pi)/2n$, but the variance of random means of $2n$ elements is $1/2n$ as is well-known. Thus it is obvious that the scatter of the stratified means is less than the scatter of the random means. Furthermore, the stratified means are distributed symmetrically since $M_3 = 0$. Observing β_2, we notice that the distribution of the stratified means is slightly more peaked than normal. Since it is well known that random means from a normal population are normally distributed, the differences between the two distributions are easy to see. As $n \to \infty$, $\beta_2 \to 3$, so it is reasonably likely that the stratified means tend to be normally distributed as the size of the sample increases.

If we select a random sample (x, y) of two elements from the normal population F, then the variance (S^2) will be:

$$S^2 = \tfrac{1}{2}(x^2 + y^2) - (x + y)^2/4 = (x - y)^2/4.$$

The method of expectations gives us the following values:

$$M_2 = \tfrac{1}{2}, \qquad M_3 = 1, \qquad M_4 = \tfrac{15}{4},$$

$$\beta_1 = 8, \qquad \beta_2 = 15, \qquad S_k = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)} = \sqrt{2}.$$

Therefore, the skewness of this distribution as measured by Elderton's formula is 1.414. For a stratified sample, where we select x from F_1 and y from F_2, the second, third, and fourth central moments of S^2 are:

$$M_2 = (\pi^2 + 2\pi + 2)/2\pi^2,$$

$$M_3 = (4\pi^3 + 7\pi^2 - 12\pi + 8)/4\pi^3,$$

$$M_4 = (15\pi^4 + 30\pi^3 - 40\pi^2 + 24\pi - 12)/4\pi^4.$$

It follows easily that $\beta_1 = 4.71084$, $\beta_2 = 10.28489$, $\kappa = 19.4$, $2\beta_2 - 3\beta_1 - 6 = .438324$, $S_k = 1.02$. For samples of two elements, the stratified samples yield a distribution for the variance which is less skewed than the corresponding distribution of the variances for random samples. The variances for random samples of two elements are distributed as a Type III curve, while the variance for stratified samples of two elements is either a Type III or a Type VI curve. The difference between the random case and the stratified case as seen from this point of view is not clear cut.

It is interesting to see what sort of bias is introduced by taking n elements of the sample from F_1 and by taking $2n$ elements of the sample from F_2. Under these circumstances, the complete sample will contain $3n$ elements, and the mean

of the sample will be $\bar{x} = \sum_{i=1}^{3n} x_i/3n = (\bar{x}_1 + 2\bar{x}_2)/3$. As before, the central moments and the β's are found to be:

$$M_2 = (\mu_{21} + 2\mu_{22})/9n = \mu_{22}/3n, \qquad M_3 = (\mu_{31} + 2\mu_{32})/27n^2 = \mu_{32}/9n^2,$$

$$M_4 = [\mu_{42} - 3\mu_{22}^2 + 9n\mu_{22}^2]/27n^3,$$

$$\beta_1 = \mu_{32}^2/3n\mu_{22}^3, \qquad \beta_2 = \mu_{42}/3n\mu_{22}^2 - 1/n + 3.$$

We notice first that the means are not symmetrically distributed for small values of n since $\beta_1 \neq 0$, but as $n \to \infty$, $\beta_1 \to 0$, so the means tend to be symmetrically distributed. It is evident also that $\beta_2 \to 3$ with increasing n; consequently, the bias which is present for small values of n tends to disappear as n increases. Incorrect proportioning of the sizes of the partial samples in stratified sampling introduces an error into the results whose magnitude decreases with an increase in n.

6. Sampling from a population $y = \phi(x)$. Suppose we have a well-behaved frequency function $\phi(x)$ of which the first four moments are finite. Furthermore, it will be required that $\phi(x)$ be continuous and Riemann-integrable. Divide the total x-axis into K parts $I_1, I_2, \cdots, I_k$ with the separating points $\alpha_1, \alpha_2, \cdots, \alpha_{k-1}$ in such manner that $\int_{-\infty}^{\alpha_1} \phi(x)\,dx = \cdots = \int_{\alpha_{k-1}}^{\infty} \phi(x)\,dx = 1/K$. In this section, we extend some of the definitions of the last section; μ_{it} will be the ith moment about the mean of the tth part I_t, and m_{it} will be the ith moment about the origin of the tth part I_t. Take a sample of Kn elements from this population so that n elements are drawn from each part. The mean of this sample will be $\bar{x} = \sum_{i=1}^{Kn} x_i/Kn$. We write this as $\bar{x} = \sum_{t=1}^{K} \bar{x}_t/K$, where $\bar{x}_t = \sum_{i=nt-n+1}^{nt} x_i/n$. It follows easily then that:

$$M_2 = \sum_{t=1}^{K} \mu_{2t}/K^2 n, \qquad M_3 = \sum_{t=1}^{K} \mu_{3t}/K^3 n^2,$$

$$M_4 = \left[\sum_{t=1}^{K} \mu_{4t} + 3n\left(\sum_{t=1}^{K} \mu_{2t}\right)^2 - 3\sum_{t=1}^{K} \mu_{2t}^2\right] \Big/ K^4 n^3,$$

as $n \to \infty$, $\beta_1 \to 0$, and $\beta_2 \to 3$. Therefore, it is evident that if we divide a population into K equal parts and take a sample of Kn elements (n elements from each part), the distribution of the means probably tends to normal as the number of elements in the sample increases.

7. Summary. Distributions of means for stratified samples have been obtained for the rectangular population which is symmetric and also for a triangular population which can be considered an example of a J-shaped population. For

both populations, the means obtained from stratified samples show less variability than the means of random samples. The stratified sample means obtained from the skewed-population exhibit less skewness than do the random sample means obtained from the same population.

The effect of stratification in sampling upon the distribution of the standard deviations is to make the distribution more symmetric. This is true for the three populations investigated.

For stratified samples from the rectangular population Student's ratio is much more stable than it is for random samples of the same size.

Thus it is evident that stratified samples possess advantages over random samples of a nature that makes stratified samples worthy of use in research work where it is easy to obtain them.

In conclusion, the author is grateful to Professor A. R. Crathorne for suggesting the problem of this paper and guiding it to its conclusion.

REFERENCES

[1] J. O. Irwin, "On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II," *Biometrika*, Vol. 19 (1927), pp. 225-239.

[2] Philip Hall, "The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable," *Biometrika*, Vol. 21 (1927), pp. 240-241.

[3] Paul R. Rider, "On the distribution of the ratio of the mean to standard deviation in small samples from non-normal universes," *Biometrika*, Vol. 21 (1929), pp. 124-143.

[4] J. Neyman, "On the two different aspects of the representative method," *Roy. Stat. Soc. Jour.*, Vol. 97 (1934), pp. 558-625.

[5] A. E. R. Church, "On the means and squared standard deviations," *Biometrika*, Vol. 18 (1926), pp. 321-394.

[6] A. A. Tchouproff, "On the mathematical expectation of the moments of frequency distributions," *Biometrika*, Vol. 12 (1918-19), pp. 140-169, 185-210.

[7] Jack Laderman, "The distribution of Student's ratio for samples of two items drawn from non-normal universes," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 376-380.

[8] W. Palin Elderton, *Frequency Curves and Correlation*, London: Charles and Edwin Layton, 1927 (Second Edition), pp. 239.

SOLUTION OF A MATHEMATICAL PROBLEM CONNECTED WITH THE THEORY OF HEREDITY[1]

By S. Bernstein

Translated by Emma Lehner

University of California, Berkeley

Translator's Note. Although a French resumé of the article here translated appeared in *Comptes Rendus*,[2] it is so condensed due to space restrictions that in reporting on Bernstein's work for the Statistical Seminar at the University of California, it became necessary to refer to the original Russian paper. Because of the obvious language difficulty together with the extreme rareness[3] of the Ukrainian publication in this country, and because of the current interest in the application of statistical theories to genetics, it seemed advisable to make this important article available to a much larger class of readers.

It is regretted that, due to the present conditions, it was impracticable to obtain the author's comments on this translation, and it is hoped that the slight changes and additions inserted, to clarify some of the more difficult passages, would have met with his approval.

---

1. Let us consider N classes of individuals which possess the property that the *cross* of any two of these individuals produces an individual belonging to one of the above N classes. We will call such a set of classes a "closed biotype." We will suppose only that the probability of obtaining an individual of class j as a result of crossing two individuals of classes i and k has some definite value $A_{ik}^j = A_{ki}^j$, and we will call these probabilities[4] "heredity coefficients of a given biotype." It follows from the definition of a closed biotype that

$$\sum_{j=1}^{N} A_{ik}^j = 1. \tag{1}$$

Let α_j be the probability that an individual belongs to class j, then under panmixia[5] the probability of belonging to class j in the next generation is given by

$$\alpha_j' = \sum_{i,k} A_{ik}^j \alpha_i \alpha_k. \tag{2}$$

[1] The original was published in the *Annales Scientifiques de l'Ukraine*, Vol. 1 (1924), p. 83–114.

[2] *C. R. Ac. Sc.*, Vol. 177, pp. 528–531, 581–584.

[3] Thanks are due to the Brown University Mathematical Library for their loan of this rare periodical.

[4] A_{ik}^j is, of course, the relative probability that an offspring belong to class j, given that the parents belong to classes i and k.

[5] That is, complete absence of selection.

Similarly we have for the next generation

$$\alpha_i'' = \sum_{i,k} A_{ik}^{i} \alpha_i' \alpha_k' \tag{3}$$

and so on.

The problem which we now propose is as follows:

For what heredity coefficients under panmixia will the distribution of probabilities achieved in the second generation remain unaltered in all subsequent generations?

If the heredity coefficients satisfy this condition, then the law of heredity which corresponds to them is called "stable."

2. We prove first of all that *the Mendelian law is stable.* The Mendelian law has to do with three classes of individuals, the first two of which are pure races, while the third is a hybrid race such that the cross of an individual of the first class with an individual of the second class *always* produces an individual of the third class. It follows therefore that

$$A_{11}^1 = A_{22}^2 = A_{12}^3 = 1, \quad \text{while}$$

$$A_{11}^2 = A_{22}^1 = A_{11}^3 = A_{22}^3 = A_{12}^1 = A_{12}^2 = 0.$$

The remaining 9 coefficients are defined as follows:

$$A_{13}^1 = A_{23}^2 = A_{13}^3 = A_{23}^3 = A_{33}^3 = 1/2$$

$$A_{33}^1 = A_{33}^2 = 1/4, \quad \text{while} \quad A_{13}^2 = A_{23}^1 = 0.$$

If, for simplicity, we denote the probabilities of belonging to the first, second or third class by α, β, γ, then (2) becomes

$$\alpha' = (\alpha + \tfrac{1}{2}\gamma)^2 \qquad \beta' = (\beta + \tfrac{1}{2}\gamma)^2 \qquad \gamma' = 2(\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma), \tag{4}$$

while on iteration we get the equivalent of (3), namely

$$\begin{aligned} \alpha'' &= [(\alpha + \tfrac{1}{2}\gamma)^2 + (\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma)]^2 = (\alpha + \tfrac{1}{2}\gamma)^2(\alpha + \beta + \gamma)^2 \\ \beta'' &= [(\beta + \tfrac{1}{2}\gamma)^2 + (\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma)]^2 = (\beta + \tfrac{1}{2}\gamma)^2(\alpha + \beta + \gamma)^2 \\ \gamma'' &= 2[(\alpha + \tfrac{1}{2}\gamma)^2 + (\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma)][(\beta + \tfrac{1}{2}\gamma)^2 + (\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma)] \\ &= 2(\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma)(\alpha + \beta + \gamma)^2. \end{aligned} \tag{5}$$

Since $\alpha + \beta + \gamma = 1$, it follows that $\alpha'' = \alpha'$, $\beta'' = \beta'$, $\gamma'' = \gamma'$, and hence the Mendelian law is stable.

3. The first rather important result can be stated as follows:

THEOREM: *If three classes form a closed biotype under a stable heredity law, which is such that the cross of an individual of the first class with an individual of the second class always produces an individual of the third class, then the first two classes represent pure races and the law of heredity is the Mendelian law.*

If the original probabilities are α, β, γ, then the corresponding probabilities for the next generation can be written as follows:

$$
\begin{aligned}
&\alpha_1 = A_{11}\alpha^2 + 2A_{12}\alpha\beta + A_{22}\beta^2 + 2A_{13}\alpha\gamma + 2A_{23}\beta\gamma + A_{33}\gamma^2 = f(\alpha, \beta, \gamma),\\
(6)\quad &\beta_1 = B_{11}\alpha^2 + 2B_{12}\alpha\beta + B_{22}\beta^2 + 2B_{13}\alpha\gamma + 2B_{23}\beta\gamma + B_{33}\gamma^2 = \varphi(\alpha, \beta, \gamma),\\
&\gamma_1 = C_{11}\alpha^2 + 2C_{12}\alpha\beta + C_{22}\beta^2 + 2C_{13}\alpha\gamma + 2C_{23}\beta\gamma + C_{33}\gamma^2 = \psi(\alpha, \beta, \gamma),
\end{aligned}
$$

where $A_{ik} + B_{ik} + C_{ik} = 1$. Since $C_{12} = 1$, by assumption, it follows that

$$B_{12} = A_{12} = 0,$$

since all the coefficients must be positive, or zero.

The mathematical problem before us consists in determining three homogeneous quadratic forms f, φ, and ψ in α, β, γ with non-negative coefficients such that

$$f + \varphi + \psi = 1 = (\alpha + \beta + \gamma)^2$$

and satisfying the conditions of stability, namely

$$
(7)\qquad
\begin{aligned}
f(\alpha_1, \beta_1, \gamma_1) &= f(\alpha, \beta, \gamma) = \alpha_1,\\
\varphi(\alpha_1, \beta_1, \gamma_1) &= \varphi(\alpha, \beta, \gamma) = \beta_1,\\
\psi(\alpha_1, \beta_1, \gamma_1) &= \psi(\alpha, \beta, \gamma) = \gamma_1,
\end{aligned}
$$

for all α, β, γ such that[6] $\alpha + \beta + \gamma = 1$. The third equation is, of course, a consequence of the first two.

The functions f, φ and ψ, being continuous, assume infinitely many values, unless they are constants, in which case they may be expressed as quadratic forms by

$$f = p(\alpha + \beta + \gamma)^2, \qquad \varphi = q(\alpha + \beta + \gamma)^2, \qquad \psi = r(\alpha + \beta + \gamma)^2$$

where p, q, r are constants. But, since the coefficient of $\alpha\beta$ is zero in f and φ, and 1 in ψ, this reduces to the trivial case $f = 0$, $\varphi = 0$ and $\psi = 1$, which we can neglect.

We now write (7) in the form

$$
(8)\qquad
\begin{aligned}
\alpha_1 &= f(\alpha, \beta, \gamma) = \alpha(\alpha + \beta + \gamma) + F_1(\alpha, \beta, \gamma)\\
\beta_1 &= \varphi(\alpha, \beta, \gamma) = \beta(\alpha + \beta + \gamma) + F_2(\alpha, \beta, \gamma)\\
\gamma_1 &= \psi(\alpha, \beta, \gamma) = \gamma(\alpha + \beta + \gamma) - F_1(\alpha, \beta, \gamma) - F_2(\alpha, \beta, \gamma).
\end{aligned}
$$

Since

$$
(9)\qquad f(\alpha, \beta, \gamma) = \alpha(\alpha + \beta + \gamma), \qquad \varphi(\alpha, \beta, \gamma) = \beta(\alpha + \beta + \gamma),
$$
$$\psi(\alpha, \beta, \gamma) = \gamma(\alpha + \beta + \gamma),$$

[6] The fact that the variables α, β, γ are not independent does not preclude the validity of identifying their coefficients in the equations that follow, since all these equations are homogeneous.

are obviously solutions of (7), it follows that $F_1(f, \varphi, \psi) = 0$ and $F_2(f, \varphi, \psi) = 0$. But, as we have just seen, f, φ and ψ assume infinitely many values. Therefore F_1 and F_2 either have a linear factor in common, or else are proportional and irreducible [7]

We first show that F_1 and F_2 do not have only a linear factor, $l\alpha + m\beta + n\gamma$, in common, for if they did this factor would vanish for $\alpha = f$, $\beta = \varphi$, $\gamma = \psi$ so that

$$(10) \qquad lf(\alpha, \beta, \gamma) + m\varphi(\alpha, \beta, \gamma) + n\psi(\alpha, \beta, \gamma) = 0.$$

But since neither f nor φ have a term in $\alpha\beta$, while ψ has, $n = 0$. Also, since f and φ have no negative coefficients, l and m are of opposite signs. Let $l \geqq 0$, while $m = -p$, where $p \geqq 0$. Then the third equation (8) can be written

$$(11) \qquad \psi(\alpha, \beta, \gamma) = \gamma(\alpha + \beta + \gamma) + (A\alpha + B\beta + C\gamma)(l\alpha - p\beta).$$

The coefficients of α^2 and β^2 in ψ must be non-negative. Therefore it follows that $A \geq 0$, while $B \leq 0$. But the coefficient of $\alpha\beta$ in ψ is 2, while $Bl - Ap$ cannot be positive. Therefore F_1 and F_2 have no linear factor in common, and must be proportional. But since the coefficient of $\alpha\beta$ in f and φ is zero, the coefficient of $\alpha\beta$ in both F_1 and F_2 must be -1, and therefore F_1 and F_2 are equal and we can write $F_1 = F_2 = F$, and (8) becomes:

$$(12) \qquad \begin{aligned} f(\alpha, \beta, \gamma) &= \alpha(\alpha + \beta + \gamma) + F(\alpha, \beta, \gamma) \\ \varphi(\alpha, \beta, \gamma) &= \beta(\alpha + \beta + \gamma) + F(\alpha, \beta, \gamma) \\ \psi(\alpha, \beta, \gamma) &= \gamma(\alpha + \beta + \gamma) - 2F(\alpha, \beta, \gamma), \end{aligned}$$

where F is an irreducible, homogeneous, quadratic form in α, β, γ. Furthermore, the coefficient of α^2 in F must be zero, since were it positive, the coefficient of α^2 in f would exceed 1, and were it negative, the coefficient of α^2 in φ would also be negative, which is impossible. Similarly, the coefficient of β^2 in F is also zero We can therefore write F in the form

$$(13) \qquad F(\alpha, \beta, \gamma) = -\alpha\beta + c\alpha\gamma + d\beta\gamma + e\gamma^2.$$

Moreover, we know that

$$(14) \qquad F(\alpha', \beta', \gamma') = F(\alpha S + F, \beta S + F, \gamma S - 2F) = 0,$$

for all values of α, β, γ, such that $\alpha + \beta + \gamma = S = 1$. Expanding (14) in Taylor series about the point $(\alpha S, \beta S, \gamma S)$ in three space we get only three terms in the expansion, since all the derivatives of order greater than the second are identically zero, and the constant term can be obtained very simply by putting $\alpha = \beta = \gamma = 0$ in $F(\alpha S + F, \beta S + F, \gamma S - 2F)$. In this way we have

[7] See Bôcher, *Introduction to Higher Algebra*, p. 210, Theorem 3.

$$(15)\quad F(\alpha S + F, \beta S + F, \gamma S - 2F)$$
$$= F(\alpha S, \beta S, \gamma S)$$
$$+ F[F'_\alpha(\alpha S, \beta S, \gamma S) + F'_\beta(\alpha S, \beta S, \gamma S) - 2F'_\gamma(\alpha S, \beta S, \gamma S)]$$
$$+ F(F, F, -2F) = 0.$$

Since F is a homogeneous form of the second degree

$$(16)\quad F(\alpha S, \beta S, \gamma S) = S^2F(\alpha, \beta, \gamma); \quad F(F, F, -2F) = F^2F(1, 1, -2),$$

while its derivatives with respect to α, β, γ are homogeneous linear forms so that

$$(17)\quad F'_\alpha(\alpha S, \beta S, \gamma S) = SF'_\alpha(\alpha, \beta, \gamma).$$

Substituting these in (15) and dividing out an $F(\alpha, \beta, \gamma)$, which is not identically zero, we get

$$(18)\quad S^2 + S[F'_\alpha(\alpha, \beta, \gamma) + F'_\beta(\alpha, \beta, \gamma) - 2F'_\gamma(\alpha, \beta, \gamma)]$$
$$= -F(\alpha, \beta, \gamma)F(1, 1, -2).$$

But since $F(\alpha, \beta, \gamma)$ is irreducible, $F(1, 1, -2)$ must be zero. Dividing by S we finally get

$$(19)\quad S = 2F'_\gamma - F'_\alpha - F'_\beta$$

or $(\alpha + \beta + \gamma) = 2(c\alpha + d\beta + 2e\gamma) - (-\beta + c\gamma) - (-\alpha + d\gamma)$ from which it follows that

$$c = d = 0, \qquad e = 1/4,$$

and hence

$$(20)\quad F(\alpha, \beta, \gamma) = \gamma^2/4 - \alpha\beta.$$

Therefore we have found that

$$f(\alpha, \beta, \gamma) = \alpha(\alpha + \beta + \gamma) + \gamma^2/4 - \alpha\beta = \alpha^2 + \alpha\gamma + \gamma^2/4 = (\alpha + \tfrac{1}{2}\gamma)^2,$$
$$(21)\quad \varphi(\alpha, \beta, \gamma) = \beta(\alpha + \beta + \gamma) + \gamma^2/4 - \alpha\beta = \beta^2 + \beta\gamma + \gamma^2/4 = (\beta + \tfrac{1}{2}\gamma)^2,$$
$$\psi(\alpha, \beta, \gamma) = \gamma(\alpha + \beta + \gamma) - \tfrac{1}{2}\gamma^2 + 2\alpha\beta = 2(\alpha + \tfrac{1}{2}\gamma)(\beta + \tfrac{1}{2}\gamma),$$

which is the Mendelian law.

4. We have therefore shown that the Mendelian law is a necessary consequence of any stable law, which is such that the cross of the first two classes produces the third hybrid class. We have not even assumed *a priori* that the first two classes are pure races. From a theoretical point of view it is interesting to investigate the possibility of crossing two pure races under different laws of heredity, which are nevertheless stable.

We will therefore suppose to start with that the coefficients of α^2 in $f(\alpha, \beta, \gamma)$

and of β^2 in $\varphi(\alpha, \beta, \gamma)$ are equal to unity. Beginning with equations (8) of the previous section, which merely express the condition that the heredity law under consideration be stable we can write

$$(22) \qquad F_1 = F = -a\alpha\beta + c\alpha\gamma + d\beta\gamma + e\gamma^2.$$

As before, F_1 and F_2 cannot have a linear factor in common, hence they are proportional, and we can write $F_2 = \lambda F$, $F_1 = F$. We therefore have five coefficients to determine: a, c, d, e, and λ. Since

$$(23) \qquad F(\alpha S + F, \beta S + \lambda F, \gamma S - (\lambda + 1)F) = 0$$

we have an analogue of (19)

$$(24) \qquad S = (1 + \lambda)F'_\gamma - F'_\alpha - \lambda F'_\beta$$

or

$$\alpha + \beta + \gamma = (1 + \lambda)(c\alpha + d\beta + 2e\gamma) + a\beta - c\gamma + \lambda a\alpha - \lambda d\gamma.$$

Equating coefficients of α, β, γ as before we have

$1 = c(1 + \lambda) + a\lambda$ or $c = (1 - \lambda a)/(1 + \lambda)$,

$1 = d(1 + \lambda) + a$ or $d = (1 - a)/(1 + \lambda)$,

$1 = 2e(1 + \lambda) - c - \lambda d = 2e(1 + \lambda) - (1 - \lambda a)/(1 + \lambda) - (1 - a)/(1 + \lambda)$,

or

$$e = (1 + \lambda - a\lambda)/(1 + \lambda)^2.$$

Therefore the most general quadratic form F satisfying our conditions can be written

$$(26) \qquad F(\alpha, \beta, \gamma) = -a\alpha\beta + \frac{1 - \lambda a}{1 + \lambda}\alpha\gamma + \frac{1 - a}{1 + \lambda}\beta\gamma + \frac{1 + \lambda - a\lambda}{(1 + \lambda)^2}\gamma^2.$$

If we let $a\lambda = b$, this becomes

$$(27) \qquad F(\alpha, \beta, \gamma) = -a\alpha\beta + \frac{1 - b}{a + b}a\alpha\gamma + \frac{1 - a}{a + b}a\beta\gamma + \frac{a + b - ab}{(a + b)^2}a\gamma^2.$$

Substituting this value of F into f, φ, ψ and simplifying, we get

$$(28) \qquad f(\alpha, \beta, \gamma) = \left(\alpha + \frac{a}{a + b}\gamma\right)\left[\alpha + (1 - a)\beta + \left(1 - \frac{ab}{a + b}\right)\gamma\right],$$

$$\varphi(\alpha, \beta, \gamma) = \left(\beta + \frac{b}{a + b}\gamma\right)\left[(1 - b)\alpha + \beta + \left(1 - \frac{ab}{a + b}\right)\gamma\right],$$

$$\psi(\alpha, \beta, \gamma) = (a + b)\left(\alpha + \frac{a}{a + b}\gamma\right)\left(\beta + \frac{b}{a + b}\gamma\right),$$

where in order that all the coefficients be positive it is necessary and sufficient that $0 \leq a \leq 1$, and $0 \leq b \leq 1$. In case $a = b = 1$ formulas (28) coincide with (21) and we get the Mendelian law.

The question of whether there actually exist heredity laws which satisfy (28) with $a < 1$, and $b < 1$ can only be solved experimentally. Theoretically formulas (28) give the most general heredity law of a closed biotype consisting of three classes, with the condition that two of the three classes be pure races. It is easy to see that the only law of heredity in which all three classes are pure races is given by the particular solution of (8)

$$(29) \qquad f = \alpha(\alpha + \beta + \gamma), \quad \varphi = \beta(\alpha + \beta + \gamma), \quad \psi = \gamma(\alpha + \beta + \gamma),$$

in which $F_1 = F_2 = 0$.

6. Supposing as before that the heredity law is stable, it remains to prove the following theorem to exhaust all possible biotypes consisting of only three classes.

THEOREM: *If all classes are hybrid, then*

$$(30) \qquad f = p(\alpha + \beta + \gamma)^2, \quad \varphi = q(\alpha + \beta + \gamma)^2, \quad \psi = r(\alpha + \beta + \gamma)^2.$$

If only one of the classes represents a pure race, then either

$$(31) \qquad \begin{aligned} f &= (\alpha + \beta)[\tfrac{1}{2}(1 + b)(\alpha + \beta) + (1 - d)\gamma] \\ \varphi &= (\alpha + \beta)[\tfrac{1}{2}(1 - b)(\alpha + \beta) + d\gamma] \\ \psi &= \gamma(\alpha + \beta + \gamma) \end{aligned}$$

or

$$(32) \qquad f = \alpha S + a\alpha(\mu\beta + \gamma) \quad \text{and} \quad \mu\varphi + \psi = 0.$$

We have seen that if f, φ, and ψ are functions of $(\alpha + \beta + \gamma)$, then we arrive at (30), in the contrary case we arrive at (8). Here we distinguish two cases: 1) F_1 and F_2 are irreducible quadratic forms which are proportional: $F_1 = k_1F$, $F_2 = k_2F$, and 2) F_1 and F_2 have a common factor, which is a linear form. Suppose at first that F is a quadratic form. If none of the numbers k_1, k_2, and $k_1 + k_2$ is zero, then two of them may be taken as positive, say k_1 and k_2. But then the coefficients of α^2 and β^2 in F would have to vanish in order that ψ have no negative coefficients. But this case of two pure races has already been discussed, and leads to formulas (28). We must therefore suppose next that one of the numbers k_1, k_2, or $k_2 + k_1$ is zero. Suppose that $k_2 + k_1 = 0$, that is, that the third class is a pure race, and hence the coefficient of γ^2 in ψ is unity. Therefore, the coefficient of γ^2 in F must be zero. We can take $k = 1$, then $k_1 = -1$, and therefore the coefficient $\alpha\gamma$ in F is negative, say $-d$. We can now write

$$F = a\alpha^2 + b\alpha\beta + c\beta^2 - d\alpha\gamma + e\beta\gamma. \tag{33}$$

We have as before

$$F(\alpha S + F, \beta S - F, \gamma S) = 0, \tag{34}$$

from which we derive by Taylor's expansion

$$S = F'_\beta - F'_\alpha \tag{35}$$

or in other words

$$\alpha + \beta + \gamma = b\alpha + 2c\beta + e\gamma - b\beta + d\gamma - 2a\alpha,$$

which leads to

$$F = \tfrac{1}{2}(b - 1)\alpha^2 + b\alpha\beta + \tfrac{1}{2}(b + 1)\beta^2 - d\alpha\gamma + (1 - d)\beta\gamma \tag{36}$$

and hence to f and φ, which are as follows,

$$f = (\alpha + \beta)[\tfrac{1}{2}(1 + b)(\alpha + \beta) + (1 - d)\gamma], \tag{37}$$

$$\varphi = (\alpha + \beta)[\tfrac{1}{2}(1 - b)(\alpha + \beta) + d\gamma].$$

It now remains to suppose that F is a linear form. Let

$$F = \lambda\alpha + \mu\beta + \gamma. \tag{38}$$

Here the condition that the heredity law be stable leads as before to the equation

$$S = (k + k_1)F'_\gamma - kF'_\alpha - k_1F'_\beta = (k + k_1) - \lambda k - \mu k_1, \tag{39}$$

where k and k_1 are linear forms

$$k = a\alpha + b\beta + c\gamma, \qquad k_1 = a_1\alpha + b_1\beta + c_1\gamma. \tag{40}$$

Hence if we had no restrictions on signs and magnitudes we could select k arbitrarily, and then we would have $k_1 = [S + (\lambda - 1)k]/[1 - \mu]$, and the solution for f, φ, ψ would depend on five parameters, (λ, μ, a, b, c).

But since in $f = \alpha S + kF$, the coefficients of β^2, $\beta\gamma$ and γ^2 are non-negative $\mu b \geq 0$, and $b + \mu c \geq 0$, $c \geq 0$, and similarly from the same property of φ we have $\lambda a_1 \geq 0$, $c_1 \geq 0$, $a_1 + \lambda c_1 \geq 0$. But μ and λ cannot both be non-negative, for then $\lambda f + \mu\varphi + \psi = 0$ would be impossible.

Let $\mu < 0$, then $b = c = 0$, but then the coefficient of α^2 in f would be $1 + a\lambda$, which will be too big, unless $\lambda = 0$. Hence, $F = \mu\beta + \gamma$, $k = a\alpha$, and

$$f = \alpha S + a\alpha(\mu\beta + \gamma), \tag{41}$$

$$\psi = -\mu\varphi = \mu[S(\beta + \gamma) - a\alpha(\mu\beta + \gamma)]/[\mu - 1].$$

Hence we have exhausted all possible cases and have proved our theorem.

6. We can summarize our results as follows. The heredity laws of a closed biotype of three classes which are stable can be divided into the following types:

1. Two classes represent pure races. The heredity laws are given by (28), and in particular for the Mendelian case by (21).

2. There are no pure races, and every race can be obtained by crossing the other races. The heredity law is given by (30).

3. All three classes are pure races. The heredity law is given by (29). Any two classes of this biotype, also form a closed biotype.

4. One of the classes is a pure race. The heredity laws are given by (31) and (32).

SOME RECENT ADVANCES IN MATHEMATICAL STATISTICS, I

By Burton H. Camp[1]

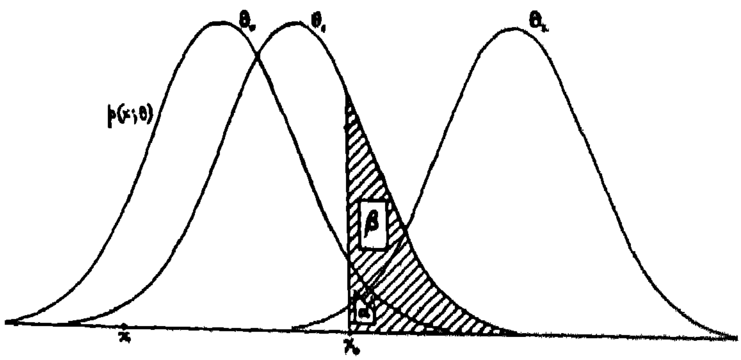
Wesleyan University

The papers considered in this partial review are listed at the end. For the most part they have appeared within the last five years, but in order to explain what has been done within the last five years it has been necessary occasionally to use material that appeared earlier. The subject matter is divided into four parts.

Part I. The Theory of Tests. Since an attempt is being made to present the material of this paper in such a form that it may be read rapidly by those who have not read the underlying literature, the author will endeavor to do little more, in Part I, than to define and illustrate several terms which are being used Altogether there are nine of these terms. It is fortunate that their meanings can be explained pretty well by reference to an extremely simple picture. Let each of the curves in the figure indicate a probability distribution $p(x; \theta)$, in which there is a single variate x and a single parameter θ.

Example 1. $p(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$, the normal distribution in which the center is at $x = \theta$, and the standard deviation is unity.

Let a random sample E be drawn from a population indicated by such a curve. In the simplest case $E = x$, a single individual. Shortly, we shall have to suppose that there are N individuals: $E = x_1, \cdots, x_N$. Eventually, the



Univariate Distributions

picture will be generalized much further. The population will be described by a function of n variables, so that, in place of each x of our sample, we shall have

[1] One of two papers read by Cecil C. Craig and by the author at a joint meeting of the Institute of Mathematical Statistics, the Econometric Society and the American Statistical Association, held in New York City on December 30, 1941.

$x^{(1)}, \cdots, x^{(n)}$; moreover there will be, not one parameter, but l parameters $\theta^{(1)}, \cdots, \theta^{(l)}$; so that our probability distribution will be multivariate and will be denoted by

$$p(x^{(1)}, \cdots, x^{(n)}; \theta^{(1)}, \cdots, \theta^{(l)}).$$

A common way of putting this is to say that x and θ are vectors in n and l dimensions, respectively, and to leave the form as originally, $p(x; \theta)$. In the figure the space which the samples ($E = x$) can occupy is of course not more than the x-axis, but in the most general case the sample space will be a part or all of a space of nN dimensions and will be denoted by W. As is well understood, a significance test is an inequality which specifies in W a certain region w as a critical region, and if E is in this w, the hypothesis being tested is rejected. For example, in the figure, one might test the hypothesis H_0 that $\theta = \theta_0$. The rejection region w_0 might be the part of the x-axis where $x > x_0$. In all such cases we shall let α equal the probability that E is in w_0 if $\theta = \theta_0$. This statement will be denoted as follows:

(1) $$\alpha = P(w_0 \mid \theta_0),$$

P standing for probability.

(*i*) *Power of a test.* A good test should satisfy two conditions: (*a*) if our sample is drawn from the population specified by θ_0, the hypothesis H_0 that $\theta = \theta_0$ should be accepted as often as possible, and (*b*) if our sample is drawn from a population specified by some other value of θ, say θ_1, then the hypothesis that $\theta = \theta_1$ should also be accepted as often as possible. Suppose first that there are but these two admissible populations. The probability of (*a*) is $1 - \alpha$. We commonly make the artificial requirement that this shall be some larger fraction such as 0.99. The probability of (*b*) is commonly denoted by β, and in the figure, when $w = w_0$, β is the area under the θ_1 curve which lies to the right of $x = x_0$. Relative to θ_0, θ_1, and α, the quantity β is called the *power* of that test which designates w_0 as the critical region. Also, α and $(1 - \beta)$ are the probabilities of the so-called errors of the first and second kinds, respectively.

(*ii*) *Unbiased test.* As stated, we would like to have β large. In any case we would like to have $\beta \gtreqqless \alpha$. If $\beta \gtreqqless \alpha$, the test and the corresponding region w_0 are "unbiased" (relative to the preassigned quantities θ_0, θ_1, and α). The region w_0 appears to be unbiased in our figure. This definition can obviously be extended to the case where, in addition to θ_1, there is an infinity of admissible values of θ; then the test is unbiased relative to the whole family of admissible values of θ if, for every one of these θ's, $\beta \gtreqqless \alpha$.

(*iii*) *UMP test and CBC region.* If, with respect to a family of admissible θ's, a critical region w_0 exists such that, for each of these θ's ($\neq \theta_0$), β is greater than it would be for any other critical region satisfying (1), then this w_0 is said to be the common best critical (CBC) region and the corresponding test is the uniformly most powerful (UMP) test.

(*iv*) *UMPU test and CBCU region.* If there is no CBC region, still it may happen that, if one restricts one's view to only unbiased regions, there may be among them a CBC region. Such a region is said to be a common best critical unbiased (CBCU) region, and the corresponding test is the uniformly most powerful unbiased (UMPU) test.

In the following examples, and elsewhere, we shall now use H_0 to indicate the hypothesis being tested, H^* to indicate all admissible alternatives.

Example 2: $p(x, \theta)$ normal as in Example 1, $E = x$, H_0: $\theta = \theta_0$, H^*: $\theta > \theta_0$. The CBC region is where $x > k$ if

$$\int_k^\infty p(x; \theta_0)\, dx = \alpha.$$

This region is the interval indicated by w_0 in the figure.

Example 3: Same as the preceding example except that now we have as H^*: $\theta \neq \theta_0$. There is no CBC region, but the CBCU region consists of two tail intervals, where $| x | > k$ if

$$\int_k^\infty p(x, \theta_0)\, dx = \tfrac{1}{2}\alpha.$$

A little reflection will convince the reader that the statements in these two examples are at least apparently true. It is geometrically evident, for example, that the last mentioned region (two tail intervals) is not as powerful with respect to the alternatives of Example 1 ($\theta > \theta_0$) as is the single tail region w_0 in the figure.

(*v*) *Type A regions.* It is often difficult to find even a CBCU region, or such a region may not exist, but it may be that there is a region which has the required properties if one admits only values of θ near to the value θ_0 being tested. Type A regions have this property. More precisely, they have the property that the power of w_0 is a minimum at θ_0 with respect to small changes in θ, and that this is a sharper minimum at θ_0 than is the power of any other w_0 which satisfies equation (1). Here the words "small changes" are used as in the calculus. The full definition [4] of an unbiased region of type A is that it shall satisfy (1) and also the following conditions:

(2) θ shall be a single parameter (not a vector),

(3) $\dfrac{d}{d\theta} P(w_0 \mid \theta) = 0$ if $\theta = \theta_0$,

(4) $\dfrac{d^2}{d\theta^2} P(w_0 \mid \theta) \geqq \dfrac{d^2}{d\theta^2} P(w \mid \theta)$ when $\theta = \theta_0$ for all regions w which satisfy the preceding conditions imposed on w_0. There are also other types of regions designated by A_1, B, C, and D, which resemble Type A [9]. The following example illustrates Type A; it is a familiar problem with an unfamiliar solution [4].

Example 4. $p(x; \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$; $E = x_1, \cdots, x_N$; $H_0: \sigma = \sigma_0$;

$H^*: \sigma \neq \sigma_0$. The UMPU region of type A is determined by two tail areas (but they are not equal tail areas) of the distribution of Σx_i^2.

(*vi*) *Test unbiased in the limit* [11]. (*vii*) *Asymptotically MP test* [15]. (*viii*) *Asymptotically MPU test* [15]. In these cases the complete definitions are too lengthy to be repeated here, and they cannot be recapitulated briefly. The general idea is that, if none of the regions of the preceding types exist, still it may be true that there are regions which do have approximately the desired properties if $E = x_1, \cdots, x_N$, and N is large. The following example [11] illustrates (*vi*).

Example 5. $p(x; \theta) = \frac{1}{\pi}\frac{1}{1 + (x - \theta)^2}$; $E = x_1, \cdots, x_N$; $H_0: \theta = 0$; $H^*: \theta \neq 0$. Regions of Type A unbiased in the limit are defined by the inequality,

$$4 \sum_i \frac{x_i^2}{(1 + x_i^2)^2} - 2 \sum_i \frac{1}{1 + x_i^2} + 4 \left(\sum \frac{x_i}{1 + x_i^2} \right)^2 \geq M \sqrt{\frac{5N}{3}} - \frac{N}{2}.$$

Here M is a quantity that has to be approximated and tabulated. The inequality is not simple, but it furnishes a definite answer to the problem.

(*ix*) *Regions similar to sample space.* All the preceding definitions apply to the case where x is a vector in n space, but not all to the case where θ is a vector in l space. Suppose now that this is the case, or, as we have said before, that there are l different parameters $\theta^{(1)}, \cdots, \theta^{(l)}$, each being capable of taking on a variety of values. Suppose we fix our attention on $\theta^{(1)}$ and wish to test the hypothesis that $\theta^{(1)} = \theta_0^{(1)}$. First of all we wish to find a critical region w_0 for which an equation like (1) will be true, independently of what the values of the other parameters may be. Such a region is said to be "similar" to sample space; the "similarity" consists in the fact that the equation like (1) would be true independently of the other parameters, if w_0 were replaced by all of sample space W, and if $\alpha = 1$. Feller [10] has shown that there are simple cases in which there is no region similar to sample space. He and others have investigated the conditions under which such regions do exist. "Generally speaking it seems that for most of the probability laws $p(x, \theta^{(1)}, \cdots, \theta^{(l)})$ in which the composite probability law for sample space is made up by multiplication,

$$\prod_{i=1}^{N} (p(x_i) \mid \theta^{(1)}, \cdots, \theta^{(l)}), \tag{2}$$

there do exist such similar regions, at least if $N > l$."

Part II. Estimation. (*i*) *Estimation by interval.* So far we have been considering possible answers to the question: Shall specified values of $\theta^{(1)}, \cdots, \theta^{(l)}$ be accepted? The totality of values of the θ's which are so acceptable might be called the acceptable point set in parameter (l-dimensional) space. This point set is determined by the sample or experiment (E), and usually different point sets are determined by different E's. Frequently this set of points consti-

tutes a simple closed region, or, in the case of only one parameter, it may be a single interval. Such an interval is called a fiducial or confidence interval. The fundamental property of such a point set or interval is well known, but has to be stated with some care: If $\alpha = 0.01$, and if one is about to take a sample from a population in which the true values of the parameters $\theta^{(1)}, \cdots, \theta^{(l)}$ are $\theta_0^{(1)}, \cdots, \theta_0^{(l)}$, then the probability is 0.99 that the sample will be such that the point set determined by it will contain this true parameter point $\theta_0^{(1)}, \cdots, \theta_0^{(l)}$. It does not matter whether or not one knows what these true values of the parameters are. If there is more than one parameter, the fiducial interval for one of these parameters often does not exist; that is, there is often no such interval which is independent of the values of the other parameters. The question whether there is such an interval is obviously connected with the question whether there are regions similar to sample space. But if one fiducial interval does exist, then usually there are an infinity of them, and our problem is to choose the best one. This problem is called "estimation by interval." One answer is to choose the shortest interval. More precisely one should say, the shortest system of intervals. One gets a system of intervals by fixing α but not E. What is desired is a formula which will give the shortest interval for every E, but it may well happen that one formula (system) will supply the shortest intervals for some E's, and another will supply the shortest intervals for other E's. The choice between the two systems will then depend on the relative frequency with which the shortest intervals will be supplied by one system or by the other.

Example 6: $p(x; \xi, \sigma)$ is normal, ξ indicating the mean and σ the standard deviation. Given $E = x_1, \cdots, x_N$; to estimate ξ. The shortest system of confidence intervals does not exist (independently of σ).

Example 7. Same as Example 6, except that now one seeks only an upper limit to the confidence interval which the parameter must not exceed. Then the shortest system (best one-sided estimate) is: $\xi \leqq \bar{x} + ts$, where Fisher's t and s are meant; t corresponds to a preassigned α, and $\bar{x}$ is the mean of the sample.

In cases like Example 6, where the shortest system does not exist, Neyman [7] defines a "short unbiased system."

Example 8. The short unbiased system for Example 6 is: $\bar{x} - ts \leqq \xi \leqq \bar{x} + ts$, $(t, s, \bar{x})$ as in Example 7.

(*ii*) *Single estimators.* Suppose that, as before, we have a sample (E) and wish to choose the best single value for one of the parameters, not as before its best fiducial interval. It is well known that there often exists a fiducial function $g(\theta)$ which, like a probability function, is everywhere positive or zero and has an integral,

$$\int_{-\infty}^{\infty} g(\theta)\, d\theta = 1,$$

and is further useful in determining confidence intervals. In particular, if θ is a location parameter and if the composite probability function is as in (2), with

only one parameter θ: $g(\theta) = kp(x_1 - \theta) \cdots p(x_N - \theta)$, k being a constant. An estimate commonly thought of as best is the maximum likelihood estimate: this is the mode of $g(\theta)$. Other estimates that have interesting properties are the mean and the median of $g(\theta)$. Pitman [14] defines a new "best" estimate θ_P. This has the property that, for every $h > 0$, θ_P is within h of the true value θ more often than is any other estimate. More precisely, if

$$P(|\theta_P - \theta| \leqq h) \geqq P(|\theta_1 - \theta| \leqq h)$$

for all positive values of h, and if the *inequality* sign between the P's holds for some positive value of h, θ_1 being every other estimate, then θ_P is the "best" estimate. As before P stands for probability.

Example 9. If $p(x; \xi, \sigma)$ is normal and the sample $E = x_1, \cdots, x_N$, the "best" estimate of σ^2 is $\dfrac{\Sigma x_i^2}{N - \frac{2}{3}}$, instead of the usual estimates: $\dfrac{\Sigma x_i^2}{N - 1}$, $\dfrac{\Sigma x_i^2}{N}$.

(*iii*) *Weight function.* Wald [13] defines a weight function $V(\theta, \theta_E)$ which depends on the seriousness of the error committed when the estimate θ_E is used in place of the true value of the parameter θ. The sample $E = x_1, \cdots, x_N$; and θ may be a vector. Thence he defines a risk function,

$$r(\theta) = \int_W V \cdot p(x_1, \cdots, x_N \mid \theta)\, dW,$$

and the "best" θ_E as that value of θ which minimizes the total risk,

$$\int Vp\, df(\theta),$$

this integral being taken over all of the parameter space, and $f(\theta)$ being the a priori distribution of θ. It is undesirable to introduce $f(\theta)$, but it can be shown that, subject to slight restrictions on the nature of f, one can obtain a best estimate by finding a value θ_E which for all θ's makes r equal to a constant and also satisfies other general conditions; this equation and these conditions do not contain $f(\theta)$. In a symmetrical but otherwise fairly general case θ_E is the maximum likelihood solution.

Part III. Likelihood Tests. This part has to do mostly with special cases of likelihood tests. As is well known, this test consists in selecting a critical rejection region w in sample space where

(*a*) $P(w \mid H_0) = \alpha$,

(*b*) the relative likelihood of H_0 is small; more precisely, where $\lambda <$ constant, and

$$\lambda = \frac{\max_\omega P(E \mid \omega)}{\max_\Omega P(E \mid \Omega)},$$

ω being the region in parameter space specified by the hypothesis tested H_0, and Ω being the region in parameter space specified by all admissible hypotheses. (In special cases *max* is replaced by *least upper bound*.) If H_0 is simple (ω being a point) and if the CBC region w exists, then w is bounded by the contour,

λ = constant [19]. Otherwise this λ test does not necessarily yield the same critical regions as do any of the preceding tests. But it is generally much easier to apply, and, in many of the cases that follow, these λ tests are good ones as judged by the preceding theory. They are powerful even if they are not the most powerful of all tests, and often this power can be found and tabulated. In fact Wilks [28] has shown that the appropriate distribution of λ (omitting terms of order $1/\sqrt{N}$) can be found[2] if the distribution of E is

$$\prod_{i=1}^{N} p(x_i, \theta^{(1)}, \cdots, \theta^{(h)}), \qquad (N \text{ large})$$

and[3] if the optimum estimates $\hat{\theta}^{(1)}, \cdots, \hat{\theta}^{(h)}$ exist and are distributed (except for certain terms of order $1/\sqrt{N}$) normally. This theorem has now been generalized by Wald, in a paper presented to the American Mathematical Society in December, 1941.

There are many of these tests, made to fit all sorts of hypotheses. The author will try to summarize a considerable group of them; all members of this group might be called generalizations of the Student-Fisher t-test. They fall naturally into two classes, according as to whether the individuals of the sample are taken from a univariate or from a multivariate universe. Unless otherwise stated all universes shall be normal. H_0 shall stand for the hypothesis being tested, and H^* for all admissible alternatives to H_0.

(i) *Univariate case.* The sample consists of N elements, as before, $x_1, \cdots, x_N$, chosen independently from N normal populations indicated by their parameters $(\xi_1, \sigma_1), \cdots, (\xi_N, \sigma_N)$. About these populations we may ask a variety of questions resulting in a variety of problems and tests.

Problem a: If the populations are all identical (ξ, σ), does $\xi = \xi_0$ (specified in advance)? This results in the well-known t-test. The hypothesis tested H_0 is that $\xi = \xi_0$, and the alternative hypothesis H^* is that $\xi \neq \xi_0$; it being assumed at the outset that all the populations are identical. The t-test has been shown to be an UMPU test relative to H^*.

Problems b, c, d: Let these same samples be arranged in k groups or "columns"

$$\begin{array}{ccc} x_1^{(1)} & \cdots & x_1^{(k)} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ x_{n_1}^{(1)} & \cdots & x_{n_k}^{(k)} \end{array}$$

where the n_i are not necessarily all equal. Let it be assumed that the populations (ξ, σ) do not change within the columns. Problems (b), (c) and (d), with their corresponding tests, may be indicated as follows:

(b) Are (ξ, σ) constant from column to column? (The $\lambda_H = L$ test.)

[2] Distribution of $(-2 \log \lambda)$ is like that of χ^2 except for terms of order $1/\sqrt{N}$.

[3] See Doob's conditions, Transactions of American Mathematical Society, vol. 36 (1934), pp 759–775.

(c) Is σ constant from column to column regardless of what values the ξ's may have? (The $\lambda_{H_1} = L_1$ test.)

(d) Is ξ constant from column to column assuming the σ's constant? (The $\lambda_{H_2} = L_2$ test.)

In Problem (b), H_0 is that (ξ, σ) are constant, H^* that they are not constant. In Problem (c), H_0 is that σ is constant, H^* that it is not constant. In Problem (d), H_0 is that ξ is constant, H^* that it is not constant. The test of Problem (c) has recently been shown to be unbiased only if the numbers in all the columns are the same ($n_1 = \cdots = n_k$). It is, however, unbiased in the limit. Power tables were published in 1937 [23]. Bartlett's (1937) μ is another test for this problem, and Pitman's [36] L test is another, but it has been shown that these two tests are equivalent. Both are unbiased; they are not likelihood tests. This problem is frequently called the problem of the "homogeneity" of a set of variances.

All these tests are, of course, functions of the observations, and the details are readily available in the papers listed. For example, Pitman's

$$L = \tfrac{1}{2}N \log \frac{\Sigma S_i}{N/2} - \Sigma\left(n_i \log \frac{S_i}{n_i/2}\right),$$

where S_i is what he calls the "squariance" for the ith column, and a large value of L is significant. The squariance is what the physicists had called and what statisticians ought therefore to have called the second moment, *viz.*: $N\mu_2$; μ_2 is really the unit second moment.

(e) *Linear Hypothesis.* Problems like the above, and many others, can be included in a general theorem by Kolodziejczyk, who showed how to write out quite simply the likelihood test if each ξ is a linear function of l parameters $(l < N)$ and if the hypothesis H_0 specifies the values of r different linear functions of the θ's $(r \leqq l)$. Furthermore, the power of this test (with numerous applications) was discussed and tabulated by Tang in an important paper [39].

Problem (f). This method (e) has been used by Neyman [43] to test the homogeneity of a set of variances, the problem already studied by a number of authors. It has been stated that some of their tests were unbiased with respect to the alternative hypothesis that the σ's were not all equal. Neyman gives reasons for supposing, in the industrial problem he is considering, that it would be more realistic to consider another alternative hypothesis, namely, H^* that the σ's are not all equal and that their distribution can be approximately described by saying that $1/\sigma^2$ has a χ^2 distribution. No UMP test exists but there does exist a critical region whose power, with respect to a sub-family of H^* is independent of the means, and the corresponding test is the most powerful test for this sub-family of alternatives. Tables of its power are furnished. More applications are promised.

(ii) *Multivariate case.* The sample consists of N elements, exactly as before, except that now each x is a vector in n space and comes from a multivariate

normal universe whose means may be represented again by ξ if we think of ξ as being a vector in n space. The other parameters of this universe are the variances and covariances α_{ij}. So, with these changes, we may repeat the statement at the beginning of (*i*) that the sample is $x_1, \cdots, x_N$, and that the populations are $(\xi_1, \alpha_{ij1}), \cdots, (\xi_N, \alpha_{ijN})$. The questions to be asked about these populations correspond exactly to those asked in the simpler case.

Problem (*a*): If the populations are all identical (ξ, α_{ij}), does $\xi = \xi_0$ (specified in advance)? The answer is given by Hotelling's T test. The hypothesis tested is H_0 that the vector $\xi = \xi_0$, and the alternative hypothesis H^* is that these two vectors are not identical. P. Hsu [28] has shown that this test is the most powerful in a special sense, and has given a new demonstration of it by the use of the Laplace transform. Incidentally he has shown that the Laplace transform of an elementary probability law determines the law uniquely except perhaps at a null set of points.

Problems (*b*), (*c*), (*d*): Now let the same sample be arranged in k groups or columns, as in (*i*) *b*, *c*, *d*; and let it be assumed that the populations (ξ, α_{ij}) do not change within the columns. Problems (*b*), (*c*), and (*d*), with their corresponding tests, may be indicated as follows:

(*b*) Are (ξ, α_{ij}) constant from column to column? (The $\lambda_{H(n)}$ test).

(*c*) Are α_{ij} constant from column to column regardless of what values the ξ's may have? (The $\lambda_{H(n')}$ test).

(*d*) Is the vector ξ constant from column to column assuming the α_{ij} constant from column to column? (The λ_H test).

Unfortunately, in the customary notation, the λ's for this case (*ii*) do not follow the pattern adopted in (*i*). It would be better to put (n) after each of the λ's (or L's) in (*i*) to signify the corresponding tests in (*ii*). But, even if this were agreed upon, there would still be a confused notation because there are many other "λ" and "L" tests besides those listed here. Apparently[4] the power functions of these last three multivariate tests have not been found yet.

(*e*) The linear hypothesis theory was shown to be applicable to the multivariate case in a special instance by P. Hsu in 1940 [38]. Since then he has generalized it further [45].

(*vii*) *Bivariate case.* This important special case of (*ii*) has now been pretty thoroughly solved. A general summary of various tests which have been devised by Finney, Pitman, Morgan, Wilks, and E. S. Pearson was given by C. Hsu in 1940 [42], with some slight additions and with tables of power functions with respect to certain alternatives. Altogether there are seven of these tests corresponding to seven different problems, including the four just referred to as Problems *a*, *b*, *c*, and *d*.

Part IV. The Method of Randomization. This part concerns randomization of the individuals within a sample to obtain a method of testing hypotheses without making use of any characteristic of the population from which the sample was drawn. It does not deal with randomization in field experi-

[4] So far as the author is aware; but he does not pretend to have made a careful search.

ments to off-set the effects of variable fertility. Also, in this discussion, the hypothesis being tested is not that the sample was a random sample. It is *assumed* that the given sample is random. We begin with an example from Pitman [46]. Two samples, $(x_1, \cdots, x_N)$ and $(y_1, \cdots, y_N)$, have been drawn at random from two populations. The means of the samples are $\bar{x}$ and $\bar{y}$, respectively. Let $|\bar{x} - \bar{y}|$ be called the spread of these samples. Now rearrange these same x's and y's with each other in all possible ways to obtain all possible spreads. The larger the observed spread, among all these possible spreads, the more significant it is supposed to be as a test of the (null) hypothesis that the two populations were identical. Similarly, tests have been devised for correlations, variances, etc.

E. S. Pearson [51] in 1938 published a criticism of this general theory which in substance seems to be that the reason why one calls the largest spreads significant, rather than the smallest ones, in the illustration just used, is that one is assuming tacitly that the admissible populations are such that large spreads would be more likely on some other than the null hypothesis; that if one does not make some such implicit assumption, then one might quite as well call the smallest spreads significant; and that therefore, barring such implicit assumptions, one can control only errors of the first kind by this method.

It seems to the author that Pearson's criticism is sound, and that, if indeed one is unwilling to make any assumption whatever about the populations considered, then this device is of no[5] value in testing the null hypothesis. For, if all that one pretends to do is to control errors of the first kind, one can do that by consulting a table of random numbers of two digits. Thus one can control errors of the first kind without performing the experiment at all, let alone making the long computations usually required by the method of randomization. Or, better, one can reduce that error to zero simply by making up one's mind that one will never reject the hypothesis being tested: certainly one will never reject it improperly if one never rejects it at all.

However, if one is willing to make in the illustration used the very mild assumption that the populations considered are such that unusually large spreads would more probably be obtained from some admissible hypothesis other than the null hypothesis, then it would seem to the author that the method would be useful. Similar remarks apply to the tests for correlations, variances, etc.

REFERENCES

PART I. THEORY OF TESTS AND PART II. ESTIMATION

[1] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc.*, A., Vol. 231 (1933), pp. 289–337.

[2] J. NEYMAN and E. S. PEARSON, "The testing of statistical hypotheses in relation to probabilities a priori," *Camb. Phil. Soc. Proc.*, Vol. 29 (1933), pp. 492–510.

[3] S. S. WILKS, "Test criteria for statistical hypotheses involving several variables," *Jour. Am. Stat. Asso.*, Vol. 30 (1935), pp. 549–560.

[4] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical

[5] Pearson's language is not so strong as this. He says "perhaps it should be described as a valuable device rather than a fundamental principle."

hypotheses I. Unbiased critical regions of type A and type A1," Stat. Res. Mem., Univ. of London, Vol. 1 (1936), pp. 1-37.

[5] J. NEYMAN and E. S. PEARSON, "Sufficient statistics and uniformly most powerful tests of statistical hypotheses," Stat. Res. Mem., Vol. 1 (1936), pp. 113-137.

[6] E. J. G. PITMAN, "The closest estimates of statistical parameters," Camb. Phil. Soc. Proc., Vol. 33 (1937), pp. 212-222.

[7] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," Phil. Trans. Roy. Soc., A, Vol. 236 (1937), pp. 333-380.

[8] J. NEYMAN, "On statistics the distribution of which are independent of the parameters involved in the probability law of the original variable," Stat. Res. Mem., Vol. 2 (1938), pp. 58-59.

[9] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," Stat. Res. Mem., Vol. 2 (1938), pp. 25-57.

[10] W. FELLER, "Note on regions similar to the sample space," Stat. Res. Mem., Vol. 2 (1938), pp. 117-125.

[11] J. NEYMAN, "Tests of statistical hypotheses which are unbiased in the limit," Annals of Math. Stat., Vol. 9 (1938), 69-86.

[12] S. S. WILKS, "Fiducial distributions in fiducial inference," Annals of Math. Stat., Vol. 9 (1938), pp. 272-280.

[13] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses," Annals of Math. Stat., Vol. 10 (1939), pp. 299-326.

[14] E. J. G. PITMAN, "The estimation of the location and scale parameters of a continuous population of any given form," Biometrika, Vol. 30 (1939), pp. 391-421.

[15] A. WALD, "Asymptotically most powerful tests of statistical hypotheses," Annals of Math. Stat., Vol. 12 (1941), pp. 1-19.

PART III LIKELIHOOD TESTS. SPECIAL CASES

[16] H. HOTELLING, "Generalized t-test," Annals of Math. Stat., Vol. 2 (1931), pp. 360-378.

[17] S. S. WILKS, "Certain generalizations in analysis of variance," Biometrika, Vol. 24 (1932), pp. 471-494.

[18] E. S. PEARSON and S. S. WILKS, "Methods of statistical analysis appropriate for k samples of two variates," Biometrika, Vol. 25 (1933), pp. 353-378.

[19] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," Phil. Trans. Roy. Soc., A., Vol. 231 (1933), pp. 289-337.

[20] B. L. WELCH, "Some problems in analysis of regression among k samples of two variables," Biometrika, Vol. 27 (1935), pp. 145-160.

[21] S. S. WILKS, "Test criteria for statistical hypotheses involving several variables," Jour. Am. Stat. Asso., Vol. 30 (1935), pp. 549-560.

[22] P. P. N. NAYER, "An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits," Stat. Res. Mem., Vol. 1 (1936), University College, London, pp. 38-56.

[23] S. S. WILKS and C. M. THOMPSON, "The sampling distribution of the criterion H when the hypothesis tested is not true," Biometrika, Vol. 29 (1937), pp. 124-132.

[24] D. J. FINNEY, "The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population," Biometrika, vol. 30 (1938), pp. 190-192.

[25] D. N. LAWLEY, "A generalization of Fisher's z-test," Biometrika, Vol. 30 (1938), pp. 180-187.

[26] P. L. HSU, "Contribution to the theory of 'Student's' t-test as applied to the problem of two samples," Stat. Res. Mem., Vol. 2 (1938), pp. 1-24.

[27] P. L. HSU, "On the best unbiased quadratic estimate of the variance," Stat. Res. Mem., Vol. 2 (1938), pp. 91-104.

[28] P. L. Hsu, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 231–243.

[29] S. S. Wilks, "The large sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 60–63.

[30] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 105–116.

[31] D. J. Bishop and U. S. Nair, "A note on certain methods of testing for the homogeneity of a set of estimated variances," *Jour. Roy. Stat. Soc., Supp.*, Vol. 6 (1939), pp. 89–99.

[32] G. W. Brown, "On the power of the L_1 test for the equality of several variances," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 119–128.

[33] E. J. G. Pitman, "A note on normal correlation," *Biometrika*, Vol. 31 (1939), pp. 9–12

[34] W. A. Morgan, "A test for the significance of the difference between the two variances in a sample from a normal bivariate population," *Biometrika*, Vol. 31 (1939), pp. 13–19.

[35] D. J. Bishop, "On a comprehensive test for the homogeneity of variances and covariances in multivariate problems," *Biometrika*, Vol. 31 (1939), pp. 31–55

[36] E. J. G. Pitman, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), pp. 200–215.

[37] P. O. Johnson and J. Neyman, "Tests of certain linear hypotheses and their application to some educational problems," *Stat. Res. Mem.*, Vol. 1 (1939), pp. 57–93

[38] P. L. Hsu, "On generalized analysis of variance, I," *Biometrika*, Vol. 31 (1940), pp. 221–237

[39] P. C. Tang, "The power function of analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 126–149.

[40] H. O. Hartley, "Testing the homogeneity of a set of variances," *Biometrika*, Vol. 31 (1940), pp. 249–255

[41] J. F. Daly, "On the unbiased character of likelihood ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 1–32.

[42] C. T. Hsu, "On samples from a normal bivariate population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 410–426.

[43] J. Neyman, "A statistical problem arising in routine analysis and in sampling inspections of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 46–76.

[44] A. Wald and R. J. Brookner, "On the distribution of Wilks' statistic for testing the independence of several groups of variates," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 137–152.

[45] P. L. Hsu, "Canonical reduction of the general regression problem," *Annals of Eugenics*, Vol. 11 (1941), pp. 42–46.

PART IV. RANDOMIZATION TESTS

[46] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (I)," *Jour. Roy. Stat. Soc. Supp.*, Vol. 4 (1937), pp. 119–130.

[47] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (II), The correlation coefficient test," *Jour. Roy. Stat. Soc. Supp.*, Vol. 4 (1937), pp. 225–232.

[48] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (III), The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322–335.

[49] B. L. Welch, "On the z-test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 26–52.

[50] B. L. Welch, "On tests for homogeneity," *Biometrika*, Vol. 30 (1938), pp. 149–158.

[51] E. S. Pearson, "Some aspects of the problem of randomization," *Biometrika*, Vol. 29 (1938), pp. 53–64.

Note: None of the 1941 *Biometrika* was received until after this paper had been read and prepared for publication.

RECENT ADVANCES IN MATHEMATICAL STATISTICS, II[1]

By Cecil C. Craig

University of Michigan

The statistical theory of the linear relationship between a dependent variable x_1, and a set of independent variables $x_2, x_3, \cdots, x_{t+1}$, is by now quite generally understood. Supposing that the x_i's are measured from their respective means, we determine the coefficients, $b_2, b_3, \cdots, b_{t+1}$, in such a way as to maximize the coefficient of correlation $r_{1.23\cdots t+1}$ between x_1 and $\sum_{i=2}^{t+1} b_i x_i$. This coefficient of correlation, usually called the multiple correlation coefficient, measures the exactness of the linear relationship that exists, and it has the property of being quite unchanged if the origins or the scales for the separate x_i's are changed in any way or even if the set $x_2, x_3, \cdots, x_{t+1}$ should be replaced by any equivalent set of linear combinations of them. That is, e.g., if $t = 3$, the new variables, $v_2 = x_2 + x_3 + x_4$, $v_3 = 2x_1 - x_3 + 3x_4$, $v_4 = x_2 + 2x_3 - 2x_4$ are equivalent to x_2, x_3, x_4, since the latter can be found if the v_i's are known, and the multiple correlation between x_1 and the v_i's is exactly the same as that between x_1 and x_2, x_3, x_4. Moreover, the requisite sampling theory if the variables involved are normally distributed is well established

I want to discuss briefly an important generalization of this kind of situation that has been the subject of recent research. In particular, in his paper, "Relations between two sets of variables," published in *Biometrika* in 1936 [1] H. Hotelling set forth these ideas in excellent fashion and contributed much to the mathematical theory required for their practical application. We now suppose that we have two sets of measurements, $x_1, \cdots, x_s$, and $x_{s+1}, \cdots, x_{s+t}$, made on the same object and that we are interested in the linear relations that may exist between the members of one set and the members of the other. As an example, $x_1, \cdots, x_s$ might be the prices of s more or less related commodities at a given time, and $x_{s+1}, \cdots, x_{s+t}$ measures of factors which may be thought to be effective in the price situation.

In the more special case I began with, $s = 1$, and a single equation fully expressed the linear statistical relationship of x_1 with $x_2, \cdots, x_{t+1}$. Now there are s dependent variables and now with $s \leqq t$, not one but s distinct linear relations will exist and will be required to fully describe the linear connections between the two sets of variables. We may assume that there is no mere duplication among the variables we are using, i.e., no one of the s x_i's is always exactly given by a linear combination of the others in the set and the same is

[1] This is the second of two papers read by B. H. Camp and the author on "Recent Advances in Mathematical Statistics" before the American Statistical Association, the Econometric Society, and the Institute of Mathematical Statistics, on December 30, 1941, in New York City The authors selected topics from papers published during the past five years.

also true of the set $x_{s+1}, \cdots, x_{s+t}$. Now there is no logical or mathematical necessity for the way in which we are so far using our measurements. Suppose $s = 2$ and $t = 3$. We can find the best linear regression equation for x_1 on x_3, x_4, x_5 and then find the like equation for x_2 on x_3, x_4, x_5. But we could very possibly get more meaning out of the situation if we began by replacing x_1 and x_2 by, say, $u_1 = x_1 + x_2$ and $u_2 = x_1 - x_2$ and similarly replacing x_3, x_4, x_5 by three v's formed from these three x's in a similar fashion. We have really been making a quite arbitrary choice among the u's and v's that could be used and the question presents itself. What significance is there in the way we choose our u's and v's?

It turns out to be much more than a merely reasonable beginning to try to determine a u from the first set and a v from the second in such a way that they will be more closely correlated than any other u and v formed in this linear fashion from the s x's in the first set and the t x's in the second. That is, we set,

$$u = \sum_{\alpha=1}^{s} a_\alpha x_\alpha \quad \text{and} \quad v = \sum_{i=s+1}^{s+t} b_i x_i,$$

and determine the a_α's and the b_i's which will maximize r_{uv}. We may say that this u and v will account for more of the linear dependence of $x_1, \cdots, x_s$ upon $x_{s+1}, \cdots, x_{s+t}$ than will any other u and v. To the mathematician familiar things begin to appear, though, as Hotelling remarks, in its purely mathematical form the problem seems to be new. A very important observation is the fact that the maximum r_{uv} would be quite unaffected by any change in origin or scale on any of the x's; it is even unaffected if we should begin by replacing the first s x's by any equivalent set of s linear combinations of them as new variables to work with and by doing the same thing on the second set of t x's. Hotelling makes use of this circumstance to greatly simplify his mathematical developments.

Now things fall out in a very interesting way. One actually solves not for the a's and b's at first but instead for the maximized r_{uv}. Having this the corresponding a's and b's can then be found. But generally the equation for r_{uv} gives not one but s different values for r_{uv}! What is the meaning of the s different r_{uv}'s? Well, you remember that I said that s relations ($s \leqq t$) would appear to exist between the two sets of variables. These s r_{uv}'s correspond to those s linear relations which are picked out in a unique way. We now have s u, v pairs which are independent of each other in the sense that no u or v is correlated with any other u or v with the exception of the other member of its pair, and of course this correlation is precisely the r_{uv} by which the pair was determined. Further, the largest r_{uv} gives the maximum u and v we set out to find; the second largest r_{uv} determines the pair u, v of maximum correlation among those independent, in the sense just described, of the first pair; the third largest r_{uv} leads to the u, v of maximum correlation among those independent of the first two pairs, and so on. The s independent linear relations among them completely describe the linear statistical dependence of the one set of variables upon the other. The relations are essentially those between the u, v pairs and

the closeness of these are measured by $r_1, r_2, \cdots, r_s$, which I write for the s r_{uv}'s. The new variables are called canonical variables and the correlations between them canonical correlations. We may say that the maximum pair, u, v, gives both the best linear predictor that can be formed from $x_{s+1}, \cdots, x_{s+t}$ and also the linear combination of $x_1, \cdots, x_s$ that can be best predicted.

I have to try to deal briefly with the numerous ideas and results in this paper which is not unrelated to earlier work by the author and by S. S. Wilks. First, what about an over all measure of the linear connection between the two sets of variables? It is shown that

$$q = \pm r_1 r_2 \cdots r_s \quad \text{and} \quad z = (1 - r_1^2)(1 - r_2^2) \cdots (1 - r_s^2),$$

have properties that make it appropriate to call the first the (vector) correlation coefficient between the two sets and the second the coefficient of alienation. Both are simply expressed by means of determinants of the covariances (product moments) among the st x's. For example, if $s = 1$, q is simply $r_{1 \cdot 23 \cdots t+1}$. If $s = t = 2$,

$$q = \frac{r_{13}r_{24} - r_{14}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{34}^2)}},$$

the numerator of which is the tetrad difference of the psychologists. Further, if it should happen that x_2 and x_4 are identical, this q becomes $r_{13 \cdot 2}$.

In an application, of course, the various quantities appearing above will have to be calculated from an observed set of values of $x_1, \cdots, x_s, x_{s+1}, \cdots, x_{s+t}$. Hotelling adapts an iterative process he had previously given to calculating the canonical $r_1, \cdots, r_s$, from which the canonical variables can be found, and he numerically illustrates the whole procedure. But what is more difficult is to solve the sampling problems that arise. It is very helpful to assume that all the x's obey a multiple normal frequency law.

First, Hotelling derives expressions for the standard errors of the r's and of q and z which are approximations useful for large samples. But for small samples exact sampling distributions are needed. Wilks [2] had earlier studied the exact sampling distribution of z in the case in which we are interested, that in the population the set $x_1, \cdots, x_s$, is completely independent of the set $x_{s+1}, \cdots, x_{s+t}$, though he did not leave his general result in a form suitable for calculation. Hotelling now finds the distribution function for q for $s = 2$. The result is not in all cases simple in form but numerical values can be obtained from it. The relations between these two possible tests, one based on z and the other based on q, are discussed at length.

An obvious undertaking would be to try to find the exact joint sampling distribution of the canonical correlations for any s and t, and I will say something about the very interesting papers in which this problem was solved. But some of this later work arose in a different though related setting which I want to discuss briefly first.

In 1936 R. A. Fisher published "The use of multiple measurements in taxonomic problems," [3] which was the introduction of linear discriminant functions

to the statistical world. Suppose that N_1 random individuals of one race (species, variety, etc.) have been measured with respect to each of k characteristics and that N_2 random individuals of another race have been similarly measured. What linear combination of these measurements would serve best to distinguish members of one race from those of the other? An example used by Fisher in this paper was that of two samples of 50 plants each of two varieties of iris found growing together in the same colony. In the flower on each plant there was measured the sepal length, x_1, the sepal width, x_2, the petal length, x_3, and the petal width, x_4. What linear function,

$$X = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_4 ,$$

would enable one to most surely identify the variety to which each single plant belongs? To choose such an X Fisher proposed the mathematical principle that the coefficients, λ_i, $i = 1, 2, 3, 4$, be determined so that the difference in the average value of X in the one variety and the average value in the other divided by the sum of squares of the X's taken about the two group means shall be a maximum. Then quite simple mathematics leads to the required numerical values of the λ_i's.

But now that we have set up such an instrument as X, there is a more interesting use to which it can be put. Suppose that the question were to establish that the N_1 individuals from the one group and the N_2 individuals from the other really belong to different races distinguishable with respect to the complex of characters we have chosen to measure in each. We are on the old question of racial likeness or unlikeness and obviously the word "race" may have a meaning broad enough to give this work of Fisher's wide application indeed. Subject to the principle according to which the coefficients λ_i are determined from sample sets of measurements, X is the best possible linear discriminant function. We are now faced with the question of the statistical significance of the difference between the means of X for each group compared to the above mentioned internal sum of squares.

It is generally useful and enlightening in a problem of this general nature turning on the use of linear and quadratic forms to consider its interpretation as an analysis of variance or covariance. Fisher readily provides such a set-up in this case by assigning to the quality of belonging to race A a numerical value, y_1, the same for all members of that race, and by assigning in like fashion a different numerical value, y_2, to the quality of belonging to race B. It is mathematically convenient if we have samples of N_1 and N_2 from races A and B respectively, to let

$$y_1 = \frac{N_2}{N_1 + N_2} \quad \text{and} \quad y_2 = -\frac{N_1}{N_1 + N_2},$$

for then over the combined sample of $N_1 + N_2$, we have,

$$S(y) = 0 \quad \text{and} \quad S(y^2) = \frac{N_1 N_2}{N_1 + N_2}.$$

This may seem somewhat arbitrary at first glance, but let us start anew by writing the linear regression equation,

$$y = \sum_{i=1}^{k} b_i(x_i - \bar{x}_i),$$

in which y takes on one of the two values above and in which $\bar{x}_i$ is the mean of x_i in the combined sample, and then proceeding to determine the b_i's in the usual least squares fashion. The b_i's turn out to be proportional to the λ_i's previously found. Now the total variance of the y's is analyzed into that within groups and that between groups and it is immediately suggested that the usual z-test with k and $N - k - 1$ degrees of freedom is the appropriate one. But, as Fisher remarks, ordinarily for the application of this test one postulates a population in which the y's have a normal distribution for each fixed set of values of $x_1, x_2, \cdots, x_k$. Here, however, the y remains fixed and one postulates a normal distribution of the x's associated with a given value of y. Not to leave this matter in doubt, though I shall return to it, I may remark that Fisher noted that earlier work by Hotelling [4] showed that the z-test is nevertheless the proper one to use.

I have to be brief indeed concerning linear discriminant functions. Fisher wrote further papers dealing with them in 1938 [5], 1939 [6], and 1940 [7] and among others, Mahalanobis [8], Bose [9, 10], and Roy [10], of the "Calcutta School" have made relevant contributions. In particular, Mahalanobis [8] introduced the concept of the generalized distance by which two sets of multiple measurements differ, which has an obvious connection with the present subject. Fisher also discussed a test for the direction in k-space in which two such samples differ most and in case we have three such samples from three different races provided a test for their collinearity.

In his 1939 paper mentioned above [6], Fisher called attention to the connections between the theory of linear discriminant functions and Hotelling's canonical correlations. Of course it can be said at once that a linear discriminant function arises as the very special case of investigating the linear relationship between the artificially introduced y and $x_1, x_2, \cdots, x_k$. And the test of significance based on the analysis of variance turns on the ratio of the sum of squares due to regression, i.e., among the predicted values, to the total sum of squares for the regression and for the residuals. This analysis is quite general in form and can equally well be set up if one is predicting linear forms formed from N_1 variables from linear forms made up from N_2 other variables. If one sets up the condition that this ratio, ϑ, be a maximum one is led, as Fisher shows, to a determinantal equation in ϑ, the roots of which are the squares of Hotelling's canonical correlations.

Mathematically the general problem we are interested in is equivalent to the following: We have a sample of $N_1 + N_2$ observed values of p normally distributed variables. If a_{ij} is the covariance of the i-th and j-th variables in the sample of N_1 and b_{ij} the like covariance in the sample of N_2 we want the sampling distribution of the roots of the determinantal equation:

$$|a_{ij} - \vartheta(a_{ij} + b_{ij})| = 0,$$

under the hypothesis that the first sample is independent of the second. This problem Fisher solved in his 1939 paper though in his characteristically concise and intuitive manner. But in the same number of the *Annals of Eugenics*, P. L. Hsu [11], at Fisher's suggestion, gave a complete analytical solution. Hsu also showed more in detail how the result applies to Hotelling's case of N observations on $s + t$ normally distributed variables in which the set of s is independent of the second set of t. In his 1936 paper Hotelling gave the result for $s = t = 2$ and in 1939, Girschick [12] gave the solution for $s = 2$ and $t > 2$. Hsu showed, too, the striking fact, mentioned by Fisher, that it is sufficient that only one of the two sets of s and of t variables be normally distributed in order that the distribution function found apply. This provides the explanation of why the test of significance applied by Fisher for linear discriminant functions is valid even though the y introduced had an arbitrary distribution of values.

The simultaneous distribution of the canonical correlations is fundamental but on finding it not all difficulties are thereby resolved. As mentioned above, either of the quantities, z or q, as they appear in Hotelling's paper, furnish over all tests, or rather they would if their distribution functions were obtained in a satisfactory form. The form of the distribution of z for complete independence was given by Wilks as early as 1932 [2] but that of q for $s > 2$ is still lacking. For $s > 2$ there are difficulties in applications even with z and in 1938 [13] M. S. Bartlett proposed a more convenient approximate test. Ordinarily, however, one would want to test the largest canonical correlation alone for significance. There are two kinds of trouble here. First, there is no assurance that the largest observed canonical correlation corresponds to the largest one in the population. Second, it is quite important to know whether the remaining population correlations are zero or not. Bartlett in 1941 [14] discussed these points.

Now I make an abrupt change in subject. Some interesting work has been done on the theory of runs and its applications during the last five years.

First, I want to try to convey some idea of the contents of three papers by W. D. Kermack and A. G. McKendrick published in 1937 [15, 16] and 1938 [17]. Suppose we have an unlimited set of numbers, no two of which are equal, and start drawing from them at random, recording the numbers in sequence as they come. Within the sequence drawn there will occur runs up and runs down of varying lengths. Thus in the sequence of 10 numbers, 2, 5, 11, 8, 9, 4, 3, 7, 14, 12, there are 3 runs up, one of length 2 and 2 of length 3, and 3 runs down, 2 of length 2 and one of length 3. Both ends of a run are counted in finding its length; no run can have a length less than 2. The total number of runs is 6 of which 3 are of length 2 and 3 are of length 3. We can also count the *gaps* which extend from crest to crest or from trough to trough and note their lengths with the convention that again both ends are counted in determining a length, so that no gap length is less than 3. Thus in the sequence of 10 numbers above there is one gap of length 3, 3 of length 4, and one of length 5.

It is clear that if we know the distribution for runs or for gaps of different

lengths we can compare an observed sequence, or rather an observed distribution of runs or gaps by lengths, with the frequencies calculated on the hypothesis of randomness and be by way of acquiring a test for this hypothesis. To be brief, in these papers these theoretical distributions are found together with their means and variances. There are some interesting applications. Tippett's random sampling numbers and a series of reversed telephone numbers both passed the χ^2-test as random and also passed the test based on the departure of the mean from its expected value compared with its standard deviation. On the other hand, the series of Swedish death rates for the period 1750-1930 could not conceivably be random. This investigation was prompted in the first place by the fluctuations of the death rate from ectromelia in mice in an experimentally induced epidemic.

The problems here dealt with had been only partially solved by earlier writers. There is much interesting material in these papers I have no space for. The authors readily include the case in which the numbers composing the population are not all different. They also studied series of limited length, series arranged in a cycle or ring and even what may be termed a Mobius cycle.

A. M. Mood in 1940 [18] in an interesting paper investigated a different form of the problem of runs. Suppose we have n elements of two kinds, say n_1 a's and $n_2 = n - n_1$ b's, and that these are arranged at random in a row. For example, if $n_1 = 5$ and $n_2 = 7$, and if a random arrangement of the 12 a's and b's is *babbbabbaaab*, the a's occur in 2 runs of one and in one run of 3 and the b's come in 2 runs of one, in one run of 2 and in one run of 3. If r_{ij} ($i = 1, 2$) is the number of runs of j of elements of variety i, Mood finds the probability of obtaining a given set of values of r_{ij} such that $\sum_j jr_{ij} = n_i$ ($i = 1, 2$), i.e., of obtaining a given pattern of runs in the two kinds of objects. Besides this basic distribution function he obtains certain marginal distributions such as that for the occurrence of a given set of runs in the a's regardless of how the b's fall (except that they must provide the necessary points of division), or that for r_1 and r_2 if these are respectively the total number of runs of a's and of b's, or that for r_1 or r_2 alone. He finds the factorial moments of these variables and then their means, variances and covariances. Similar results are obtained in case there are more than two kinds of elements. In the second part of the paper, Mood turns to the case of drawings from an infinite population in which articles of two or more kinds occur in fixed proportions. Finally, in both of the two kinds of drawings considered he derives the limiting forms of the distributions studied as the sample size increases. As Mood notes, here, too, a few of the results had previously been found, but this paper is the first really thoroughgoing investigation of its subject.

In a paper antedating Mood's by some six months, A. Wald and J. Wolfowitz [19] used the distribution function for the total number of runs (irrespective of length) for arrangements of fixed numbers of two kinds of elements to provide a test of the hypothesis that two samples have come from the same population with a continuous distribution law. If the observations in the two samples

combined are arranged in order of magnitude and if then the observations from the first sample are each replaced by a zero and those from the second are each replaced by a one, we have a situation to which this distribution function for runs applies. W. L. Stevens in 1939 [20] also discussed an application of this distribution.

The third principal topic I have chosen for my remarks is developments in the use of the probability integral transformation. The use of this device at all seems to be quite recent, appearing in a paper by H. Cramer in 1928 [21] who invented a test of goodness of fit which reappeared as the "ω^2-test" in apparently independent work of R. von Mises in 1931 [22]. In 1932 in a section new in the fourth edition of "Statistical Methods for Research Workers," [23] Fisher showed the usefulness of this transformation in combining independent tests of significance and in 1933 and 1934 Karl Pearson [24, 25] had papers in *Biometrika* on the subject.

As for the transformation itself, suppose that $p(x)$ is the probability density function of a continuous variable x defined on the range (a, b) such that,

$$\int_a^b p(x)\,dx = 1.$$

Then let us introduce the variable,

$$y = \int_a^x p(x)\,dx,$$

which is the probability that a value of the variable at random will be less than x. It will be seen that since x is a random variable, the proportion of population values less than an x drawn at random is itself a random variable. Perhaps this will be clearer if I use a simple example of J. Neyman's to show how a sample of x's also determines a sample of y's for a given $p(x)$. Suppose that,

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

and that a sample of 5 values of x arranged in order of magnitude is: -1.5, -1.1, -0.5, 0.6, 1.6. Then by reference to a table of areas under the normal curve of error, we find that the corresponding observed y's are: 0.067, 0.136, 0.309, 0.726, 0.945. It is obvious that the range for y is always, for any $p(x)$, $(0, 1)$. Further if $f(y)$ is the probability density function for y, of course,

$$f(y)\,dy = p(x)\,dx.$$

But from the definition of y,

$$dy = p(x)\,dx,$$

so that $f(y) = 1$. Thus, quite independently of $p(x)$, y obeys a rectangular distribution law on the range $(0, 1)$.

This simplicity of the distribution of the quantity y and its independence of

$p(x)$ are most attractive properties. I shall note briefly some of the applications that have been made in recent years.

In 1936 W. R. Thompson [26] denoted by p_k the probability that in a sample of N a randomly chosen x will be less than x_k, the k-th value observed. Then the probability that $p' \leqq p_k \leqq p''$ is just $p'' - p'$. The probability that exactly r other members of the sample will be less than x_k is then,

$$\binom{N-1}{r} p_k^r (1 - p_k)^{N-1-r}.$$

Further for all samples in which just r values occur less than x_k, the proportion of occasions on which $p' \leqq p_k \leqq p''$ is given by

$$\int_{p'}^{p''} p^r (1 - p)^{N-r-1} \, dp \Big/ \beta(r + 1, N - r),$$

the difference of two incomplete β-functions. But that there are exactly r observed x's less than x_k is equivalent to saying that x_k is the $(r + 1)$-st observation in order of magnitude, so that in the above we may as well replace r by $k - 1$. It is easy to find that the expected value of p_k in such samples is $\dfrac{k}{N+1}$ and that the variance is $\dfrac{k(N - k + 1)}{(N + 1)^2(N + 2)}$. It follows from the first of these two expressions that the proportion of occasions on which $x_k < x < x_{N-k+1}$ is $\dfrac{N + 1 - 2k}{N + 1}$, $(N + 1 > 2k)$. Statements of this kind establish confidence limits. Thus if one says that in a sample of N, an observation at random will fall between the k-th and the $(N - k + 1)$-st observations in order of magnitude, such a statement has a probability of $\dfrac{N + 1 - 2k}{N + 1}$ of being true. Or, the integral just above is the fiducial probability of the truth of $p' \leqq p_k \leqq p''$ if in a sample of N the k-th observation is the $(r + 1)$-st in order of magnitude. Thompson went on to obtain confidence limits for the median in a sample of N from any population.

In 1939 Wald and Wolfowitz [27] studied the problem of obtaining confidence limits for $\varphi(x)$, the proportion of observations in a sample of N with values less than a given x, the population obeying any continuous distribution law. Their arguments are too complicated to attempt to sketch them here, but they are based on the fact that the transformed variable, y, as defined above, is rectangularly distributed on the interval $(0, 1)$. With their exact solution they gave a more convenient approximate method for calculation in applications.

In 1938 (I am not being strictly chronological) E. S. Pearson [28] published a study of test criteria based on this probability integral transformation. Suppose that we have n independently observed y's, $y_1, y_2, \cdots, y_n$. How should the y's be used to test the hypothesis that the observations from which the y's were calculated all came from the same population? K. Pearson [24] had

already suggested the use of $Q = y_1y_2 \cdots y_n$ or $Q' = (1 - y_1)(1 - y_2) \cdots (1 - y_n)$. It is known that a simple function of Q or of Q' obeys a χ^2-distribution with $2n$ degrees of freedom so that we have a ready means of combining independent tests based on Q or Q'. But how is one to choose among Q, Q', or other functions of the y's that might be suggested? E. S. Pearson emphasized the role that the hypotheses conceived as alternate to the one being tested should play in making such a choice. He illustrates this in a case of testing the hypothesis that a sample came from a normal population of zero mean and unit variance and in which the alternate populations, from one of which the sample might have been drawn, are such that the corresponding y's calculated on the hypothesis being tested would follow a Pearson type I distribution law. Using the likelihood principle he was led in this case to Q or Q', which are then concluded to be "best possible tests."

The final paper I want to discuss is an important one by J. Neyman on the "Smooth test of goodness of fit," published in 1937 [29]. Suppose again that a random sample of N values of x gives the set, $y_1, y_2, \cdots y_N$ on the hypothesis H_0 that the population distribution law is $p(x \mid H_0)$. If H_0 is true the y's in random samples do follow a rectangular distribution on (0, 1). But what would be the distribution of the y's if the distribution law for the population were actually $p(x \mid H_1)$? We have for the y's as calculated,

$$y = \int_a^x p(x \mid H_0)\, dx.$$

But to find $f(y)$,

$$f(y)\, dy = p(x \mid H_1)\, dx,$$

so that,

$$f(y) = \frac{p(x \mid H_1)}{p(x \mid H_0)} \neq 1.$$

Therefore if H_0 is not true, the y's calculated on the assumption that it is may be expected to exhibit a statistically significant set of deviations from a rectangular distribution.

As Neyman remarks, it is a defect of the χ^2-test of goodness of fit that the information one has of the algebraic signs of the differences between calculated and observed frequencies, particularly of the way in which positive and negative differences succeed each other, is completely unused. And in forming a test of a statistical hypothesis it is now well understood, thanks to Neyman and Pearson, that due account should be taken of the alternate hypotheses conceivably true.

Neyman begins by specifying a wide class of alternate hypotheses in a form that lends itself to mathematical treatment. This is done by assuming that the distribution of y's calculated for H_0 will, if an alternate H_1 is true, be given by a function of the form,

$$p(y \mid \theta_1, \theta_2, \cdots, \theta_k) = ce^{\sum_1^k \theta_i \pi_i(y)}$$

in which $\pi_i(y)$ is a polynomial of degree i (a transformed Legendre polynomial) with convenient properties. For low values of k, such as will ordinarily be used, this permits alternate distribution curves to deviate in a smooth manner from the distribution tested, with a limited number of intersections with it.

Now the problem is to determine the function of the observed y's which will provide a suitable test of H_0 with respect to the alternate hypotheses of order or class k, k having been decided upon in advance of making the test. The mathematics, proceeding along Neyman and Pearson lines, shows that the appropriate function, for large samples at least, is simply $\sum_1^k u_i^2$ in which,

$$u_i = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \pi_i(y_j')$$

the y_j''s being calculated from the sample. Moreover, the probability that the sum $\sum_1^k u_i^2$ exceeds a given value is at once obtained from a table of the incomplete Γ-function, i.e., this sum is proportional to a χ^2.

This is a very fine piece of work but, as Neyman points out, there are still questions to be settled concerning the general utility of this "smooth test." F. N. David in 1939 [30] further discussed this test. In particular, it may be pointed out that the parameters in $p(x \mid H_0)$ must be assumed known; what would be the effect on the test of estimating these parameters is unknown. A reasonably large sample seems to be required to make the developments on the assumption of large samples applicable but a y must be calculated for each observation. This makes for a good deal of computing but it is not known how grouping of observations might be effected. And the matter of the choice of the order of the test to be applied, i.e., of a value of k, is still somewhat in doubt.

I will not debate the proposition that there are papers completely omitted from this discussion as important as those I have included however inadequately. The limitations of space forced me to choose and it is quite possible that my personal tastes and interests had more weight than they should.

REFERENCES

[1] H. Hotelling, "Relations between two sets of variates," *Biometrika*, 28 (1936), 323–377.

[2] S. S. Wilks, "Certain generalisations in the analysis of variance," *Biometrika*, 24 (1932), 471–494.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7 (1936), 179–188.

[4] H. Hotelling, "The generalization of Student's ratio," *Annals of Math. Stat.*, 2 (1931), 360–378.

[5] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, 8 (1938), 376–386.

[6] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, 9 (1939), 238–249.

[7] R. A. Fisher, "The precision of discriminant functions," *Annals of Eugenics*, 10 (1940), 422–429.

[8] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. Ind.*, 12 (1936), 49-55.

[9] R. C. Bose, "On the exact distribution of the D^2 statistic," *Sankhya*, 2 (1936), 143-154.

[10] R. C. Bose and S. N. Roy, "The exact distribution of the Studentized D^2 statistic," *Sankhya*, 3 (1938), part 4.

[11] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, 9 (1939), 250-258.

[12] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Annals of Math. Stat.*, 10 (1939), 203-224.

[13] M. S. Bartlett, "Further aspects of the theory of multiple regression," *Proc. Camb. Phil. Soc.*, 34 (1938), 33-40.

[14] M. S. Bartlett, "The statistical significance of canonical correlations," *Biometrika*, 32 (1941), 29-37.

[15] W. D. Kermack and A. G. McKendrick, "Tests for randomness in a series of numerical observations," *Roy. Soc. Edin. Proc.*, 57 (1937), 228-240.

[16] W. D. Kermack and A. G. McKendrick, "Some distributions associated with a randomly arranged set of numbers," *Ibid.*, 332-376.

[17] W. D. Kermack and A. G. McKendrick, "Some properties of points arranged at random on a Mobius surface," *Mathematical Gazette*, 22 (1938), 66-72.

[18] A. M. Mood, "The distribution theory of runs," *Annals of Math. Stat.*, 11 (1940), 367-392.

[19] A. Wald and J. Wolfowitz, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, 11 (1940), 147-162.

[20] W. L. Stevens, "Distribution of groups in a sequence of alternatives," *Annals of Eugenics*, 9 (1939), 10-17.

[21] H. Cramér, "On the composition of elementary errors, Second paper, Statistical applications," *Skandinavisk Aktuarietidskrift*, 11 (1928), 141-180.

[22] R. von Mises, "Vorlesungen aus dem Gebiete der angewandten Matematik," Bd. 1: *Wahrscheinlichkeitsrechnung*, (Leipzig, 1931), 316-335.

[23] R. A. Fisher, *Statistical Methods for Research Workers*, (Edinburgh, 4th edition, 1932), Article 21.1.

[24] K. Pearson, "On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random," *Biometrika*, 25 (1933), 379-410.

[25] K. Pearson, "On a new method of determining 'goodness of fit,'" *Biometrika*, 26 (1934), 425-442.

[26] W. R. Thompson, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math. Stat.*, 7 (1936), 122-128.

[27] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, 10 (1939), 105-118.

[28] E. S. Pearson, "The probability integral transformation for testing goodness of fit and combining independent tests of significance." *Biometrika*, 30 (1938), 134-148.

[29] J. Neyman, "A smooth test of goodness of fit," *Skandinavisk Aktuarietidskrift*, 20 (1937), 149-199.

[30] F. N. David, "On Neyman's 'smooth' test of goodness of fit. I. Distribution of the criterion ψ^2 when the hypothesis tested is true," *Biometrika*, 31 (1939), 191-199.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A FURTHER REMARK CONCERNING THE DISTRIBUTION OF THE RATIO OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE TO THE VARIANCE[1]

By John von Neumann

Institute for Advanced Study[2]

1 Introduction. In our previous paper[1] it was found convenient to assume that the number m (of the variables of the quadratic form under consideration) is even. (Cf. p. 383, loc. cit.) This means that in the application to the mean square successive difference $n = m + 1$ must be odd. (Cf. p. 389, id.)

In this note we shall show that the distribution for an odd m (i.e. an even n) can be expressed by means of the distribution for an even m, the latter being already known, loc. cit.

Specifically, consider the distribution of $\gamma = \sum_{\mu=1}^{m} a_\mu x_\mu^2$, if the $x_1, \cdots, x_m$ are equidistributed over the surface $\sum_{\mu=1}^{m} x_\mu^2 = 1$. Denote the m-uplet $(a_1, \cdots, a_m)$ by A, then the distribution function of γ depends on A; denote that distribution by $\omega_A(\gamma)$. (Cf. p. 372 id., we write a_μ for the B_μ there.)

Now consider an m-uplet $A = (a_1, \cdots, a_m)$ and a p-uplet $B = (b_1, \cdots, b_p)$ and form the $m + p$-uplet $C = (a_1, \cdots, a_m, b_1, \cdots, b_p)$. Write $C = A + B$. Then we shall show that there exists a simple expression for $\omega_C(\gamma)$ in terms of $\omega_A(\gamma)$ and $\omega_B(\gamma)$.

For the specific application to the mean square successive difference, we can put $n = m + 1$, $A = (\cos(\pi\mu/n)$ for $\mu = 1, \cdots, \frac{1}{2}n - 1, \frac{1}{2}n + 1, \cdots, n - 1)$, $B = (0)$, $C = A + B = (\cos \pi\mu/n$ for $\mu = 1, \cdots, n - 1)$.

2. The recursion formula. We proceed as follows. $\omega_A(\gamma)$ can also be used to express the joint statistics of

$$\gamma = \sum_{\mu=1}^{m} a_\mu x_\mu^2 \quad \text{and} \quad \rho = \sum_{\mu=1}^{m} x_\mu^2,$$

or better, the volume of that part of the $x_1, \cdots, x_m$-space which corresponds to any given domain in the γ, ρ-plane. Thus the volume corresponding to a

[1] Cf. the paper by the same author, *Annals of Math. Stat.*, vol. 12(1941), pp. 367–395.

[2] Also Scientific Advisory Committee of the Ballistic Research Laboratory, Aberdeen Proving Ground.

given infinitesimal γ, ρ domain $d\gamma\, d\rho$ will clearly be

$$C_m \sqrt{\rho}^{m-1}\, d\sqrt{\rho}\cdot \omega_A\left(\frac{\gamma}{\rho}\right)\frac{d\gamma}{\rho},$$

where C_m is the $(m-1$-dimensional) area of the $x_1, \cdots, x_m$-surface $\sum_{\mu=1}^{m} x_\mu^2 = 1$, (the unit sphere). I.e., this volume is

$$(1) \qquad \tfrac{1}{2} C_m \omega_A\left(\frac{\gamma}{\rho}\right)\cdot \rho^{\frac{1}{2}m-2}\, d\gamma\, d\rho.$$

Similarly for

$$\zeta = \sum_{\nu=1}^{p} b_\nu u_\nu^2 \quad \text{and} \quad \sigma = \sum_{\nu=1}^{p} u_\nu^2$$

the volume corresponding to the infinitesimal ζ, σ domain $d\zeta\, d\sigma$ is

$$(2) \qquad \tfrac{1}{2} C_p \omega_B\left(\frac{\zeta}{\sigma}\right)\cdot \sigma^{\frac{1}{2}p-2}\, d\zeta\, d\sigma.$$

Finally for $\theta = \gamma + \zeta = \sum_{\mu=1}^{m} a_\mu x_\mu^2 + \sum_{\nu=1}^{p} b_\nu u_\nu^2$ and $\tau = \rho + \sigma = \sum_{\mu=1}^{m} x_\mu^2 + \sum_{\nu=1}^{p} u^2$ the volume cosresponding to the infinitesimal θ, τ domain $d\theta\, d\tau$ is

$$(3) \qquad \tfrac{1}{2} C_{m+p}\, \omega_{A+B}\left(\frac{\theta}{\tau}\right)\cdot \tau^{\frac{1}{2}(m+p)-2}\, d\theta\, d\tau.$$

Now $\theta = \gamma + \zeta$, $\tau = \rho + \sigma$ connect (1), (2), (3) as follows:

$$\tfrac{1}{2} C_{m+p}\, \omega_{A+B}\left(\frac{\theta}{\tau}\right)\tau^{\frac{1}{2}(m+p)-2}$$
$$= \int_0^\tau d\rho\cdot\int d\gamma\cdot \tfrac{1}{2} C_m \omega_A\left(\frac{\gamma}{\rho}\right)\rho^{\frac{1}{2}m-2}\cdot \tfrac{1}{2} C_p \omega_B\left(\frac{\theta-\gamma}{\tau-\rho}\right)(\tau-\rho)^{\frac{1}{2}p-2}.$$

This gives (either by simply putting $\tau = 1$, or else by replacing θ, γ, ρ by $\tau\theta$, $\tau\gamma$, $\tau\rho$)

$$\omega_{A+B}(\theta) = \frac{C_m C_p}{2C_{m+p}} \int_0^1 d\rho\cdot\int d\gamma\cdot \omega_A\left(\frac{\gamma}{\rho}\right)\omega_B\left(\frac{\theta-\gamma}{1-\rho}\right)\rho^{\frac{1}{2}m-2}(1-\rho)^{\frac{1}{2}p-2}.$$

To determine $\dfrac{C_m C_p}{2C_{m+p}}$ apply to this $\int d\theta \cdots$. Then

$$1 = \frac{C_m C_p}{2C_{m+p}} \int_0^1 d\rho\cdot \rho^{\frac{1}{2}m-1}(1-\rho)^{\frac{1}{2}p-1}$$
$$= \frac{C_m C_p}{2C_{m+p}} B[\tfrac{1}{2}m, \tfrac{1}{2}p] = \frac{C_m C_p}{2C_{m+p}} \frac{\Gamma[\frac{1}{2}m]\Gamma[\frac{1}{2}p]}{\Gamma[\frac{1}{2}(m+p)]}.$$

Accordingly:

$$(\mathrm{I}) \qquad \omega_{A+B}(\theta) = \frac{\Gamma[\frac{1}{2}(m+p)]}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}p)} \int_0^1 d\rho\cdot\int d\gamma\cdot\omega\left(\frac{\gamma}{\rho}\right)\omega\left(\frac{\theta-\gamma}{1-\rho}\right)\rho^{\frac{1}{2}m-2}(1-\rho)^{\frac{1}{2}p-2}.$$

3. The special case. Let us now return to the special case mentioned at the end of 1—the application to the mean square successive difference.

There $p = 1$ and $B = (0)$, so that the "distribution" of ξ is concentrated at the point 0. Hence $\omega_B(\zeta)$ is an "improper" distribution, concentrated in the same way.[3] Using C and A as described at the end of 1, the above formula becomes (now $m = n - 2$, $p = 1$)

$$\text{(II)} \qquad \omega_{A+(0)}(\theta) = \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}(n-2)]\Gamma[\frac{1}{2}]} \int_0^1 d\rho \cdot \omega_A\left(\frac{\theta}{\rho}\right) \rho^{\frac{1}{2}m-1}(1-\rho)^{-\frac{1}{2}}$$

It would have been equally easy, of course, to establish (II) directly.

Putting $\rho = 1/t$ gives

$$\text{(III)} \qquad \omega_{A+(0)}(\theta) = \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}(n-2)]\Gamma[\frac{1}{2}]} \int_1^\infty dt \cdot \omega_A(\theta t) t^{-\frac{1}{2}m-\frac{1}{2}}(t-1)^{-\frac{1}{2}}.$$

Since $\omega_A(\gamma)$ vanishes for $|\gamma| > \cos(\pi/n)$, we may replace the integral $\int_1^\infty$ by $\int_1^{\cos(\pi/n)/|\theta|}$.

Formula (III) can be used for numerical work, and also to extend the formula (3) on p. 391, loc. cit., to even values of n.

CONVEXITY PROPERTIES OF GENERALIZED MEAN VALUE FUNCTIONS

By E. F. Beckenbach

University of Michigan

In an article appearing in the *Annals of Mathematical Statistics*[1] it was pointed out that while the mean value functions appearing below have been studied and used since 1840, there appeared to have been no attempt made to investigate the behavior of their second derivatives.

Consider (1) the unit weight or simple sample form

$$\varphi(t) = \left(\frac{x_1^t + x_2^t + \cdots + x_n^t}{n}\right)^{1/t},$$

in which the x_i are positive numbers and in which t may take any real value; (2) the weighted sample form

$$\omega(t) = \left(\frac{c_1 x_1^t + c_2 x_2^t + \cdots + c_n x_n^t}{c_1 + c_2 + \cdots + c_n}\right)^{1/t},$$

---

[3] Dirac's famous "delta function." It could be described by a Stieltjes Integral.

[1] Nilan Norris, "Convexity properties of generalized mean value functions," *Annals of Math Stat.*, Vol. 8 (1937), pp. 118-120.

in which the c_i are positive numbers, and in which the x_i and t are restricted as in $\varphi(t)$; and (3) the integral form

$$\theta(t) = \left(\frac{1}{x_2 - x_1} \int_{x_1}^{x_2} [f(x)]^t \, dx \right)^{1/t},$$

in which $f(x)$ is a positive continuous function for $x_1 \leqq x \leqq x_2$.

Since the analysis and results are essentially the same in all three cases, we restrict our attention to $\theta(t)$

As is well known,[2] $\theta(t)$ is a monotone non-decreasing function which varies from the minimum of $f(x)$ to the maximum of $f(x)$ as t increases from $-\infty$ to $+\infty$. It is further of some importance to study the rate at which the rate of increase of this type bias is changing as t increases; the rate in question is given by the second derivative $\theta''(t)$.

The following points were made by Norris, loc. cit.: (1) Since, as we have pointed out, $\theta(t)$ has two horizontal asymptotes, $\theta(t)$ must have at least one inflection point. (2) Consideration of a simple example shows that there is not necessarily an inflection point at $t = 0$; $\theta''(0)$ can be made to take on any real value.

Thus it is *not* true that $\theta''(t)$ must be positive for all $t < 0$ and negative for all $t > 0$. On the other hand, we shall give simple bounds for $\theta''(t)$ in the other direction; namely, we shall give a positive upper bound of $\theta''(t)$ for $t < 0$ and a lower bound for $t > 0$. These bounds are precise in the sense that they are actually taken on in the special case $f(x) \equiv$ const. Their main advantage lies in the fact that while the expression for $\theta''(t)$ is quite involved, these bounds are simple expressions in the quantities $\theta(t)$ and $\theta'(t)$ which might already have been computed.

Let

$$\lambda(t) \equiv \log \theta(t).$$

Differentiating, we obtain

$$t^2\lambda'(t) = t^2 \frac{\theta'(t)}{\theta(t)} = \frac{\int_{x_1}^{x_2} [f(x)]^t \log [f(x)]^t \, dx}{\int_{x_1}^{x_2} [f(x)]^t \, dx} - \log \left(\frac{1}{x_2 - x_1} \int_{x_1}^{x_2} [f(x)]^t \, dx \right).$$

It follows[3] that

$$\lambda'(t) \geqq 0$$

and

$$\theta'(t) \geqq 0.$$

Let

$$\mu(t) \equiv t^2\lambda'(t).$$

[2] See for instance, G. Pólya und G. Szegö, *Aufgaben und Lehrsätze aus der Analysis* (Berlin, 1925), Vol. 1, pp. 54–55 and 210–211.

[3] See G. Pólya und G. Szegö, loc. cit., p. 210.

Curiously, while $\lambda''(t)$ and $\theta''(t)$ appear to be rather formidable, the closely related quantity $\mu'(t)$ is made relatively simple by the fact that two of the terms obtained by formal differentiation are negatives of each other; and Schwarz' inequality can be applied to the remaining terms, as follows.

We obtain

$$\left(\int_{x_1}^{x_2} [f(x)]^t\, dx\right)^2 \mu'(t) = t\left[\left(\int_{x_1}^{x_2} [f(x)]^t\, dx\right)\left(\int_{x_1}^{x_2} [f(x)]^t [\log f(x)]^2\, dx\right) - \left(\int_{x_1}^{x_2} [f(x)]^t \log f(x)\, dx\right)^2\right].$$

By Schwarz' inequality,[4] it follows that

$$\mu'(t) = t\pi(t),$$

with

$$\pi(t) \geqq 0,$$

the sign of equality holding if and only if $f(x) \equiv$ const.

From the definition of $\mu(t)$ we obtain

$$\mu'(t) = t[2\lambda'(t) + t\lambda''(t)] = \frac{t}{\theta(t)}\left[2\theta'(t) + t\theta''(t) - \frac{t[\theta'(t)]^2}{\theta(t)}\right],$$

whence

$$2\lambda'(t) + t\lambda''(t) = \frac{1}{\theta(t)}\left[2\theta'(t) + t\theta''(t) - \frac{t[\theta'(t)]^2}{\theta(t)}\right] = \pi(t) \geqq 0;$$

that is,

$$t\lambda''(t) \geqq -2\lambda'(t), \qquad t\theta''(t) \geqq \frac{t[\theta'(t)]^2}{\theta(t)} - 2\theta'(t).$$

It follows that for $t < 0$, we have

$$\lambda''(t) \leqq -2\lambda'(t)/t$$

and

$$\theta''(t) \leqq \frac{[\theta'(t)]^2}{\theta(t)} - \frac{2\theta'(t)}{t};$$

while for $t > 0$, we have

$$\lambda''(t) \geqq -2\lambda'(t)/t$$

and

$$\theta''(t) \geqq \frac{[\theta'(t)]^2}{\theta(t)} - \frac{2\theta'(t)}{t}.$$

[4] See G. Pólya und G. Szegö, loc. cit., p. 54.

A CHARACTERIZATION OF THE NORMAL DISTRIBUTION

By Eugene Lukacs

Baltimore, Md.

1. In sampling from a normal population the distributions of the mean and of the variance are mutually independent. This well known property of the normal distribution is used in deriving the distribution of "Student's" ratio. The independence of the distributions of the mean and of the variance characterizes the normal distribution. To show this one has to prove the following statement:

A necessary and sufficient condition for the normality of the parent distribution is that the sampling distributions of the mean and of the variance be independent.

That this condition is necessary follows from the above mentioned property of the normal distribution; so there is only to prove that this condition is sufficient. This was first proved by R. C. Geary[1] by using some of R. A. Fisher's general formulae for the seminvariants. However, a different proof, using characteristic functions might be of some interest.

2. Let $f(x)$ be the density function of a continuous probability distribution and let $x_1, x_2, \cdots, x_n$ be n observations of the variate x. Denote by $\bar{x} = \sum_{\alpha=1}^{n} x_\alpha/n$ the sample mean, and by

$$s^2 = \sum_{\alpha=1}^{n} (x_\alpha - \bar{x})^2/n = [(n-1)\sum_{1}^{n} x_\alpha^2 - 2\sum_{\alpha=1}^{n-1}\sum_{\beta=\alpha}^{n-1} x_\alpha x_{\beta+1}]/n^2$$

the sample variance of these observations. The characteristic function of the distribution is then given by

$$\psi(t) = \int e^{itx} f(x)\,dx. \tag{1}$$

The characteristic function of the joint distribution of the statistics $\bar{x}$ and s^2 is known to be

$$\varphi(t_1, t_2) = \int \cdots \int e^{it_1\bar{x} + it_2 s^2} f(x_1) \cdots f(x_n)\,dx_1 \cdots dx_n. \tag{2}$$

In the same way one obtains the characteristic function of the mean $\bar{x}$ as

$$\varphi_1(t_1) = \varphi(t_1, 0) = \int \cdots \int e^{it_1\bar{x}} f(x_1) \cdots f(x_n)\,dx_1 \cdots dx_n, \tag{2a}$$

[1] R. C. Geary, "Distribution of Student's ratio for nonnormal samples," *Roy. Stat. Soc. Jour.*, Supp. Vol. 3, no. 2.

and the characteristic function of the distribution of the variance s^2

$$\varphi_2(t_2) = \varphi(0, t_2) = \int \cdots \int e^{it_2 s^2} f(x_1) \cdots f(x_n)\, dx_1 \cdots dx_n. \tag{2b}$$

The independence of the distributions of $\bar{x}$ and s^2 means in terms of the characteristic functions $\varphi(t_1, t_2) = \varphi_1(t_1)\varphi_2(t_2)$, or

$$\left.\frac{\partial \varphi(t_1, t_2)}{\partial t_2}\right|_{t_2=0} = \varphi_1(t_1) \left.\frac{\partial \varphi_2(t_2)}{\partial t_2}\right|_{t_2=0}. \tag{3}$$

Substituting in (2a) $\bar{x} = \sum_1^n x_\alpha \Big/ n$, it is seen

$$\varphi_1(t_1) = \prod_{\alpha=1}^{n} \int e^{it_1 x_\alpha/n} f(x_\alpha)\, dx_\alpha = \left[\int e^{it_1 x/n} f(x)\, dx\right]^n = [\psi(t_1/n)]^n,$$

therefore

$$\left.\frac{\partial \varphi}{\partial t_2}\right|_{t_2=0} = [\psi(t_1/n)]^n \left.\frac{\partial \varphi_2}{\partial t_2}\right|_{t_2=0}. \tag{3'}$$

Differentiating (2) with respect to t_2

$$\frac{\partial \varphi(t_1 t_2)}{\partial t_2} = i \int \cdots \int s^2 e^{it_1 \bar{x} + it_2 s^2} f(x_1) \cdots f(x_n)\, dx_1 \cdots dx_n$$

Substituting $s^2 = [(n-1)\sum x_\alpha^2 - 2\sum\sum x_\alpha x_\beta]/n^2$ and $\bar{x} = \sum_1^n x_\alpha/n$ we obtain easily

$$\left.\frac{\partial \varphi(t_1, t_2)}{\partial t_2}\right|_{t_2=0} = \frac{(n-1)i}{n}\left\{[\psi(t_1/n)]^{n-1} \int x^2 e^{it_1 x/n} f(x)\, dx - [\psi(t_1/n)]^{n-2}\left[\int x e^{it_1 x/n} f(x)\, dx\right]^2\right\}. \tag{4a}$$

In a similar way it is seen

$$\left.\frac{\partial \varphi_2(t_2)}{\partial t_2}\right|_{t_2=0} = \frac{i(n-1)}{n}\sigma^2. \tag{4b}$$

Here σ^2 denotes the population variance of the parent distribution. Substituting (4a) and (4b) in the relation (3′) and writing $t = t_1/n$ one has

$$\psi(t) \int x^2 e^{itx} f(x)\, dx - \left[\int x e^{itx} f(x)\, dx\right]^2 = [\psi(t)]^2 \sigma^2. \tag{5}$$

Considering the definition (1) of the characteristic function it is seen that

$$\frac{d^l \psi(t)}{dt^l} = i^l \int x^l e^{itx} f(x)\, dx. \tag{6}$$

The integrals on the left side of relation (5) are of this form. So one may write the relation expressing statistical independence of the sample mean and the sample variance as a differential equation for the characteristic function $\psi(t)$, namely

$$-\psi(t)\frac{d^2\psi}{dt^2} + \left(\frac{d\psi}{dt}\right)^2 = \sigma^2[\psi(t)]^2. \tag{7}$$

The initial conditions to be satisfied are

$$\psi(0) = 1, \qquad \psi'(0) = i\mu, \tag{7a}$$

where μ is the population mean of the parent distribution. Integrating this equation it is seen that the characteristic function is

$$\psi(t) = e^{i\mu t}e^{-\frac{1}{2}\sigma^2 t^2}, \tag{8}$$

which is the characteristic function of the normal distribution.

3. This reasoning applies also to the multivariate case. Let $f(x_1, x_2, \cdots, x_p)$ be the density of the p variates $x_1, x_2, \cdots, x_p$. Denote by $x_{k\alpha}$ ($k = 1, 2, \cdots, p$; $\alpha = 1, 2, \cdots, n$) the α-th observation on the k-th variate, by $\bar{x}_k$ the sample mean of this variate and by s_{kl} the sample covariance between the k-th and l-th variates. Assuming that the distribution of s_{kl} is independent of the joint distribution of the p sample means $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)$ one obtains the equation

$$\frac{\psi_{lm}}{\psi} - \frac{\psi_l\psi_m}{\psi^2} = -\sigma_{lm}. \tag{9}$$

Here σ_{lm} is the population covariance of the variates x_l and x_m,

$$\psi = \psi(t_1, \cdots t_p) = \int \cdots \int e^{i(t_1x_1+\cdots+t_px_p)} f(x_1, \cdots, x_p)\, dx_1 \cdots dx_p,$$

denotes the characteristic function of the parent distribution and

$$\psi_l = \frac{\partial\psi}{\partial t_l}, \qquad \psi_{lm} = \frac{\partial^2\psi}{\partial t_l\,\partial t_m}.$$

If (9) holds for $l, m = 1, 2, \cdots, p$ one has a system of partial differential equations which leads to the characteristic function of the multivariate normal distribution.

NOTE ON A METHOD OF SAMPLING

By Carlos E. Dieulefait

National University of Litoral, Argentina

Olds[1] has considered the following problem: *Given a lot of size $m = s + r$ containing s items of a specified kind. Items are drawn without replacement until j of the s items have been drawn. The problem is to determine the probability law of n, the number of drawings which have to be made.* In the present note, we shall consider a certain limiting form for the probability function of n and make some remarks concerning repeated sampling of this type.

If n is the size of a drawing $j \leq n \leq r + j$ its probability law $P(n)$ is given by:

$$P(n) = \frac{C_{r,n-j}C_{s,j}}{C_{m,n}}\cdot\frac{j}{n} = \frac{\Gamma(s+1)}{\Gamma(j)\Gamma(s-j+1)}\,C_{r,n-j}\int_0^1 x^{n-1}(1-x)^{m-n}\,dx.$$

The characteristic function of n is

$$\varphi(t, n) = \sum_{n=j}^{r+j} P(n)e^{nt} =$$

$$\frac{\Gamma(s+1)}{\Gamma(j)\Gamma(s-j+1)}\cdot e^{jt}\int_0^1 x^{j-1}(1-x)^{s-j}(1-x+xe^t)^r\,dx.$$

Differentiating we find

$$\text{(1)}\qquad \frac{\varphi'(t, n)}{\varphi(t, n)} = j + re^t\,\frac{\int_0^1 x^j(1-x)^{s-j}(1-x+xe^t)^{r-1}\,dx}{\int_0^1 x^{j-1}(1-x)^{s-j}(1-x+xe^t)^r\,dx}$$

and hence

$$m_1(n) = \sum_{n=j}^{r+j} P(n)n = [\varphi'(t, n)]_{t=0} = j\cdot\frac{m+1}{s+1}.$$

For the calculation of moments about the mean we take

$$\text{(2)}\qquad \varphi(t, n - m_1) = e^{-m_1 t}\varphi(t, n),$$

from which we obtain

$$[\varphi^{(k)}(t, n - m_1)]_{t=0} = \sum_{n=j}^{r+j} P(n)(n - m_1)^k = \mu_k(n).$$

In particular, $\mu_2 = \dfrac{rj(m+1)(s+1-j)}{(s+1)^2(s+2)}$. The values of $m_1(n)$ and $\mu_2(n)$ have already been given by Olds using another method. Putting $\dfrac{rj}{s+1} = \beta$, we have

[1] E. G Olds, *Annals of Math. Stat.*, Vol. 11 (1940), p. 355.

$$\mu_2 = \beta(1 - \beta) + \frac{r^{(2)}(j, 2)}{(s + 1, 2)}$$

$$\mu_3 = 3(1 - \beta)\mu_2 + \beta(2 - 3\beta + \beta^2) + \frac{r^{(3)}(j, 3)}{(s + 1, 3)}$$

$$\mu_4 = (6 - 4\beta)\mu_3 - (11 + 4\beta + 6\beta^2)\mu_2 - \beta(6 - 11\beta + 6\beta^2 + \beta^3) + \frac{r^{(4)}(j, 4)}{(s + 1, 4)},$$

where $r^{(k)} = r(r - 1) \cdots (r - k + 1)$, $(j, k) = j(j + 1) \cdots (j + k - 1)$.[2]

We can obtain a limiting form for $P(n)$ in the following way: Since

$$\varphi\left(t, \frac{n - j}{r}\right) = e^{-jt/r} \varphi\left(\frac{t}{r}, n\right)$$

we find

$$\varphi\left(t, \frac{n - j}{r}\right) = \frac{\Gamma(s + 1)}{\Gamma(j)\Gamma(s - j + 1)} \int_0^1 x^{j-1}(1 - x)^{s-j}(1 - x + xe^{t/r})^r \, dx.$$

Therefore

$$\lim_{r \to \infty} \varphi\left(t, \frac{n - j}{r}\right) = \int_0^1 L(x)e^{xt} \, dx, \tag{3}$$

where

$$L(x) = \frac{\Gamma(s + 1)}{\Gamma(j)\Gamma(s - j + 1)} x^{j-1}(1 - x)^{s-j}.$$

The interpretation of (3) is that the distribution $\left\{\frac{n - j}{r};\ P(n)\right\}$ has as its limiting form the distribution $\{x, L(x)\}$ as $r \to \infty$.

Letting $n_1, n_2, \cdots n_w$ be a sample of size w and $\bar{n}$ the mean, $\bar{n} = \frac{1}{w} \sum_{i=1}^{w} n_i$. For the characteristic function of $\bar{n}$ we have

$$\varphi(t, \bar{n}) = \sum_{n_i = j}^{r+j} \prod_{i=1}^{w} P(n_i)e^{n_i t/w} = \prod_{i=1}^{w} \varphi\left(\frac{t}{w}, n\right) = \left[\varphi\left(\frac{t}{w}, n\right)\right]^w$$

and hence

$$\frac{\varphi'(t, \bar{n})}{\varphi(t, \bar{n})} = w \frac{\varphi'\left(\frac{t}{w}, n\right)}{\varphi\left(\frac{t}{w}, n\right)} = \left[\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t = t/w}.$$

[2] For an easy symbolical method of calculation cf. C. Dieulefait, *Comptes Rendu*, Vol. 208, p. 145.

For $t = 0$ we have $m_1(\bar{n}) = m_1(n)$. But:

$$\frac{d^\alpha}{dt^\alpha}\frac{\varphi'(t, \bar{n})}{\varphi(t, \bar{n})} = \frac{1}{w^\alpha}\left[\frac{d^\alpha}{dt^\alpha}\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t=t/w}$$

Then for $t = 0$ we arrive at

$$\mu_{\alpha+1}(\bar{n}) = \frac{\mu_{\alpha+1}(n)}{w^\alpha}.$$

For $\alpha = 1$, we have

$$\mu_2(\bar{n}) = \frac{\mu_2(n)}{w}$$

and this leads us to

$$\sigma_{\bar{n}} = \sqrt{\frac{rj(m+1)(s+1-j)}{w(s+1)^2(s+2)}}.$$

By the Tchebycheff theorem we obtain

$$P(|\bar{n} - m_1(n)| < l\sigma_{\bar{n}}) > 1 - \frac{1}{l^2}.$$

We can take l and w as large as we please; then we have the following stochastic limit

$$\lim_{w\to\infty} \bar{n} = m_1(n).$$

Now, we have

$$\Phi_w(t) = \varphi\left(t, \frac{\bar{n} - m_1}{\sigma_{\bar{n}}}\right) = e^{-m_1 t/\sigma_{\bar{n}}}\left[\varphi\left(\frac{t}{\sigma_{\bar{n}} w}, n\right)\right]^w$$

and

$$\frac{\Phi_w'(t)}{\Phi_w(t)} = -\frac{m_1}{\sigma_{\bar{n}}} + \frac{1}{\sigma_{\bar{n}}}\left[\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t=t/\sigma_{\bar{n}} w}.$$

Remembering (1) we readily obtain

$$\frac{\Phi_w'(t)}{\Phi_w(t)} = \frac{\frac{rt}{\sigma_{\bar{n}}^2 w}\left[-\frac{rj^2}{(s+1)^2} + \frac{(r-1)j(j+1)}{(s+1)(s+2)} + \frac{j}{s+1}\right] + \cdots}{1 + \frac{rt}{\sigma_{\bar{n}} w}\frac{j}{s+1} + \cdots}$$

$$= \frac{t + \cdots}{1 + \frac{rt}{\sigma_{\bar{n}} w}\frac{j}{s+1} + \cdots}.$$

Thus, we find

$$\lim_{w\to\infty} \frac{\Phi_w'(t)}{\Phi_w(t)} = t.$$

This result implies that the distribution $\left\{\frac{\bar{n}_w - m_1(\bar{n})}{\sigma_{\bar{n}}}; P(\bar{n})\right\}$ has the limiting normal distribution $\left\{x, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right\}$, as $w \to \infty$.

A SEQUENCE OF DISCRETE VARIABLES EXHIBITING CORRELATION DUE TO COMMON ELEMENTS

By Carl H. Fischer

University of Michigan

1. Introduction. Studies of correlation due to common elements have been made more or less sporadically over the past thirty years in attempts to throw more light on the meaning of correlation. Numerous examples may be cited. One of the earliest was a study by Kapteyn [1] in which he showed that two sums, each of n elements drawn from a normal population with k elements in common, had a correlation coefficient of k/n. This was considerably generalized by the writer [3] who considered sums of different numbers of elements drawn from quite arbitrary continuous distributions. The work was extended to include sequences of three or more such sums. Antedating this latter paper, Rietz [2] has devised various urn schemata in one of which pairs of drawings of s balls each were produced with t balls held in common. The coefficient of correlation between the numbers of white balls in each of the pairs of drawings was found to be t/s.

Fairly recently some interest has been shown in this subject in connection with the study of heredity; hence it appeared that it might be of value to present the following study by elementary methods of a sequence of discrete variables in which each member is linked to the adjacent members by various specified numbers of common elements.

2. Two variables. A pair of discrete variables is defined as follows: The first, x, is equal to the number of white balls in a set of s_1 balls drawn one at a time from an urn which is so maintained that the probability of drawing a white ball is always a constant, p. The second, y, is equal to the number of white balls in a second set of s_2 balls formed by drawing t_{12} balls at random from the s_1 balls of the first set plus $s_2 - t_{12}$ balls drawn directly from the urn. The numbers s_1 and s_2 may or may not be equal.

Evidently the marginal distribution of x follows the Bernoulli law and is given by $\binom{s_1}{x} q^{s_1-x} p^x$.[1] The first step in finding $P(x, y:t_{12})$, the bivariate distribution

[1] By $\binom{a}{b}$ is meant the number of combinations of a items taken b at a time. It shall be understood that $\binom{a}{b} = 0$ if $b < 0$ or $b > a$.

function of x and y with t_{12} balls in common between the two drawings, is to write the product of the three probabilities: of obtaining x white balls in the first set; of drawing d of these whites in the t_{12} balls chosen at random from this set; of drawing exactly $y - d$ white balls among the $s_2 - t_{12}$ balls drawn directly from the urn to complete the second set. This product may readily be reduced to the form shown below in (1), symmetric in x and y and in s_1 and s_2, which is then summed on d from 0 to t_{12}. Thus

$$(1) \qquad P(x, y : t_{12}) = \sum_{d=0}^{t_{12}} \binom{s_1 - t_{12}}{x - d}\binom{t_{12}}{d}\binom{s_2 - t_{12}}{y - d} q^{s_1+s_2-t_{12}-x-y+d} p^{x+y-d}.$$

The marginal distribution of x has already been given. From the symmetry of (1) it is obvious that the corresponding marginal distribution of y must be characterized by the Bernoulli distribution function $\binom{s_2}{y} q^{s_2-y}p^y$. The variances of the marginal distributions are s_1pq and s_2pq, respectively.

We next proceed to demonstrate that both of the regression curves are linear and to find the equations of the lines. Consider an array of x on y for some fixed value of y. The mean of the array is

$$(2) \qquad \bar{x}_y = \binom{s_2}{y}^{-1} q^{-(s_2-y)} p^{-y} \sum_{x=0}^{s_1} xP(x, y : t_{12}).$$

The summation in the right member of (2) may be expanded and then rewritten as

$$(3) \qquad \sum_{d=0}^{t_{12}} \binom{t_{12}}{d}\binom{s_2 - t_{12}}{y - d} q^{s_2-y}p^y \sum_{x=0}^{s_1} x \binom{s_1 - t_{12}}{x - d} q^{s_1-t_{12}-x+d} p^{x-d}.$$

The inner summation in (3) is seen to equal $d + p(s_1 - t_{12})$ and hence (2) becomes

$$\begin{aligned} \bar{x}_y &= \binom{s_2}{y}^{-1}\left\{\sum_{d=0}^{t_{12}} \binom{t_{12}}{d}\binom{s_2 - t_{12}}{y - d} [d + p(s_1 - t_{12})]\right\} \\ &= \binom{s_2}{y}^{-1}\left\{t_{12} \sum_{d=0}^{t_{12}} \binom{t_{12} - 1}{d - 1}\binom{s_2 - t_{12}}{y - d} + p(s_1 - t_{12})\binom{s_2}{y}\right\}. \end{aligned}$$

Then the equation of the line of regression of x on y becomes

$$(4) \qquad \bar{x}_y = t_{12}y/s_2 + p(s_1 - t_{12}).$$

By symmetry, the line of regression of y on x may be seen to be

$$\bar{y}_x = t_{12}x/s_1 + p(s_2 - t_{12}).$$

The square of the correlation coefficient is equal to the product of the slopes of the two regression lines, hence

$$(5) \qquad r_{xy} = t_{12}/(s_1s_2)^{\frac{1}{2}}.$$

If $s_1 = s_2 = s$ we have the familiar result t/s.

3. **Three variables.** A third variable, z, may now be defined as the number of white balls in a set of s_3 balls formed by drawing t_{23} balls at random from the s_2 of the second set plus $s_3 - t_{23}$ drawn directly from the urn. It is evident from the results on two variables that the marginal distribution of z follows the Bernoulli law and that the equations of the regression lines of z on y and y on z are

$$\bar{z}_y = t_{23}y/s_2 + p(s_3 - t_{23});$$
$$\bar{y}_z = t_{23}z/s_3 + p(s_2 - t_{23}).$$

The correlation coefficient, r_{yz}, is equal to $t_{23}/(s_2s_3)^{\frac{1}{2}}$.

The relationship between x and z remains to be investigated. The probability of the joint occurrence of x whites on the first drawing and z whites on the third when it is specified that the s_1 and s_3 balls of the two sets shall include the same g balls in common is given by the right member of (1) with g, z, and s_3 replacing t_{12}, y, and s_2, respectively. When this expression is multiplied by the probability that the first and third sets do contain exactly g balls in common and the product is summed on g over the range 0 to t_{12}, we have $P(x, z:t_{12}, t_{23})$, the bivariate distribution function of x and z. Thus

$$(6) \qquad P(x, z:t_{12}, t_{23}) = \sum_{g=0}^{t_{12}} \binom{t_{12}}{g}\binom{s_2 - t_{12}}{t_{23} - g}\binom{s_2}{t_{23}}^{-1} P(x, z:g).$$

The mean of the array of x and z for any fixed z may be written, after inverting the order of summation:

$$(7) \quad \bar{x}_z = \sum_{g=0}^{t_{12}} \left\{\left[\binom{s_3}{z}^{-1} q^{-(s_3-z)} p^{-z} \sum_{x=0}^{s_1} xP(x, z:g)\right]\binom{t_{12}}{g}\binom{s_2 - t_{12}}{t_{23} - g}\binom{s_2}{t_{23}}^{-1}\right\}.$$

The expression within the square brackets of (7) is identical in form with the right member of (2), and hence we now have

$$\bar{x}_z = \sum_{g=0}^{t_{12}} \left\{[gz/s_3 + p(s_1 - g)]\binom{t_{12}}{g}\binom{s_2 - t_{12}}{t_{23} - g}\binom{s_2}{t_{23}}^{-1}\right\}.$$

This reduces readily to

$$(8) \qquad \bar{x}_z = \frac{t_{12}t_{23}}{s_2 s_3} z + \frac{s_1 s_2 - t_{12}t_{23}}{s_2} p.$$

By symmetry,

$$\bar{z}_x = \frac{t_{12}t_{23}}{s_1 s_2} x + \frac{s_2 s_3 - t_{12}t_{23}}{s_2} p.$$

The coefficient of correlation between x and z is found to be

$$(9) \qquad r_{xz} = \frac{t_{12}t_{23}}{s_2(s_1 s_3)^{\frac{1}{2}}}.$$

It will be observed that

$$(10) \qquad r_{xz} = r_{xy} r_{yz} .$$

Interesting relationships also exist among the partial and multiple correlation coefficients and the multiple regression surfaces. It will be convenient here to measure each variate from its mean and to replace the subscripts x, y, and z, on r by 1, 2, and 3, respectively. Then the multiple regression surface of each variable on the other two may be conveniently expressed in terms of the cofactors of the correlation determinant. From the results found by the writer [4] for the case where each element r_{ij} of the correlation determinant may be expressed as the product $r_{i,i+1} \cdot r_{i+1,i+2} \cdots r_{j-1,j}$, we now have

$$R_{11} = 1 - r_{23}^2 , \qquad R_{12} = -r_{12}(1 - r_{23}^2) ,$$

$$R_{22} = 1 - r_{13}^2 , \qquad R_{23} = -r_{23}(1 - r_{12}^2) ,$$

$$R_{33} = 1 - r_{12}^2 , \qquad R_{13} = 0 .$$

Then the regression planes of x on y and z and of z on x and y are given, respectively, by

$$x = \frac{r_{12}\sigma_1}{\sigma_2} y = \frac{t_{12}}{s_2} y ,$$

$$z = \frac{r_{23}\sigma_3}{\sigma_2} y = \frac{t_{23}}{s_2} y .$$

The regression plane of y on x and z is

$$y = \frac{\sigma_2}{1 - r_{12}^2 r_{23}^2} \left\{ \frac{r_{12}(1 - r_{23}^2)}{\sigma_1} x + \frac{r_{23}(1 - r_{12}^2)}{\sigma_3} z \right\}$$

$$= \frac{(s_2^2 s_3 - s_2 t_{23}^2) t_{12}}{s_1 s_2^2 s_3 - t_{12} t_{23}} x + \frac{(s_1 s_2^2 - s_2 t_{12}^2) t_{23}}{s_1 s_2^2 s_3 - t_{12} t_{23}} z .$$

The three multiple correlation coefficients are

$$(11) \qquad r_{1\cdot 23} = r_{12} , \qquad r_{3\cdot 12} = r_{23} , \qquad r_{2\cdot 13} = \left[\frac{1 - (1 - r_{12}^2)(1 - r_{23}^2)}{1 - r_{12}^2 r_{23}^2} \right]^{\frac{1}{2}}$$

The partial correlation coefficients are

$$(12) \quad r_{12\cdot 3} = r_{12} \left[\frac{1 - r_{23}^2}{1 - r_{12}^2 r_{23}^2} \right]^{\frac{1}{2}} , \qquad r_{23\cdot 1} = r_{23} \left[\frac{1 - r_{12}^2}{1 - r_{12}^2 r_{23}^2} \right]^{\frac{1}{2}} , \qquad r_{13\cdot 2} = 0$$

4. **k variables.** A sequence of k variables may be formed successively as were the three considered above. It will be convenient here to designate the variables by x_i ($i = 1, 2, \cdots k$). We also define h_i as the total number of balls held in common between the first and the i-th drawings. Then, as special cases, $h_1 = s_1$ and $h_2 = t_{12}$.

The bivariate distribution functions, regression lines, and correlation coefficients associated with any two consecutive variables in the sequence and with any two variables separated by only one other variable can, from the preceding results, be written at once.

It is not difficult to derive the bivariate distribution function for x_1 and x_k by an extension of the method used in deriving (6). We then have

$$
\begin{aligned}
&P(x_1, x_k : t_{12}, t_{23} \cdots t_{k-1,k}) \\
(13) \qquad &= \sum_{h_k} \sum_{h_{k-1}} \cdots \sum_{h_3} \left\{ \prod_{i=3}^{k} \left[\binom{h_{i-1}}{h_i} \binom{s_{i-1} - h_{i-1}}{t_{i-1,i} - h_i} \binom{s_{i-1}}{t_{i-1,i}}^{-1} \right] P(x_1, x_k : h_k) \right\}.
\end{aligned}
$$

The equation of the line of regression of x_1 on x_k is

$$\bar{x}_1 = \sum_{x_1=0}^{s_1} x_1 P(x_1, x_k : t_{12}, t_{23} \cdots t_{k-1,k}).$$

This may be reduced, by repeated applications of the steps illustrated in the corresponding case for three variable, to the form

$$(14) \qquad \bar{x}_1 = \frac{t_{12} t_{23} \cdots t_{k-1,k}}{s_2 s_3 \cdots s_k} x_k + \frac{s_1 s_2 \cdots s_{k-1} - t_{12} t_{23} \cdots t_{k-1,k}}{s_2 s_3 \cdots s_{k-1}} p.$$

By symmetry, we have

$$\bar{x}_k = \frac{t_{12} t_{23} \cdots t_{k-1,k}}{s_1 s_2 \cdots s_{k-1}} x_1 + \frac{s_2 s_3 \cdots s_k - t_{12} t_{23} \cdots t_{k-1,k}}{s_2 s_3 \cdots s_{k-1}} p.$$

Then the simple correlation coefficient between x_1 and x_k is

$$(15) \qquad r_{1k} = \frac{t_{12} t_{23} \cdots t_{k-1,k}}{s_2 s_3 \cdots s_{k-1} (s_1 s_k)^{1/2}} = r_{12} \cdot r_{23} \cdots r_{k-1,k}.$$

It was shown by the writer [4] that for a sequence such as we are considering the multiple correlation coefficient is a function only of the variables immediately adjacent to the one considered, and that the partial correlation coefficient is zero for any pairs except those of consecutive variables in the sequence. Thus, the formulas given in terms of simple correlation coefficients for the case of a sequence of three variables may be interpreted so as to cover the case for k variables.

REFERENCES

[1] J. C. KAPTEYN, "Definition of the correlation-coefficient," *Monthly Notices Roy. Astron Soc.*, Vol. 72(1912), pp. 518–525.

[2] H. L. RIETZ, "Urn schemata as a basis for the development of correlation theory," *Annals of Math.* Vol. 21(1920), pp. 306–322.

[3] C. H FISCHER, "On correlation surfaces of sums with a certain number of random elements in common," *Annals of Math. Stat.* Vol. 4(1933), pp. 103–126.

[4] C. H. FISCHER, "On multiple and partial correlation coefficients of a certain sequence of sums," *Annals of Math. Stat.* Vol. 4(1933), pp. 278–284.

REPORT OF THE NEW YORK MEETING OF THE INSTITUTE

The Seventh Annual Meeting of the Institute of Mathematical Statistics was held from Saturday to Tuesday, December 27-30, 1941, in conjunction with the meetings of the Allied Social Science Associations. With the exception of the session on Tuesday afternoon, all sessions were held at the Biltmore Hotel. The following one hundred seventy-seven members* of the Institute attended the meeting:

F. L. Alt, H. E. Arnold, K. J. Arnold, L. A. Aroian, K. J. Arrow, R. W. Bachelor, I. L. Battin, B. M. Bennett, Carl Bennett, Joseph Berkson, Felix Bernstein, F. E. Blanche, C. I. Bliss, A. J. Bonis, Paul Boschan, A. H. Bowker, D. S. Brady, A. E. Brandt, R. H. Brown, R. W. Burgess, J. H. Bushey, Belle Calderon, B. H. Camp, J. M. Clarkson, W. G. Cochran, A. C. Cohen, Jr., M. S. Cohen, Isadore Cohn, J. B. Coleman, L. M. Court, D. R. G. Cowan, Gertrude Cox, C. C. Craig, B. B. Day, D. B. DeLury, W. E. Deming, W. J. Dixon, H. F. Dodge, H. F. Dorn, Paul Dorweiler, David Durand, J. H. Dutka, P. S. Dwyer, Churchill Eisenhart, W. F. Elkin, J. S. Elston, M. L. Elveback, D. R. Embody, W. D. Evans, Willy Feller, J. W. Fertig, Irving Fisher, W. C. Flaherty, M. M. Flood, R. M. Foster, L. K. Frankel, H. A. Freeman, G. R. Gause, Hilda Geiringer, C. H. Graves, J. A. Greenwood, J. I. Griffin, C. C. Grove, F. E. Grubbs, E. J. Gumbel, M. J. Hagood, H. J. Hand, M. H. Hansen, Myron Heidingsfield, Edward Helly, G. M. Hopper, Harold Hotelling, L. A. Hoy, William Hurwitz, Seymour Jablon, W. W. Jacobs, Rachel Jenss, Myron Kantorowitz, Karl Karsten, Leo Katz, C. J. Kiernan, B. F. Kimball, A. J. King, L. F. Knudsen, H. S. Konijn, Tjalling Koopmans, R. L. Kozelka, A. K. Kurtz, A. R. Kury, S. M. Kwerel, Jack Laderman, Oscar Lange, D. H. Leavens, B. A. Lengyel, Howard Levene, Ida Levin, M. J. Lass, Irving Lorge, A. J. Lotka, Eugene Lukacs, G. A. Lundberg, P. J. McCarthy, W. G. Madow, Benjamin Malzberg, Henry Mann, Jakob Marschak, J. W. Mauchly, G. F. T. Mayer, Margaret Merrell, J. N. Michie, J. R. Miner, Nathan Morrison, J. E. Morton, F. C. Mosteller, M. R. Neifeld, Harold Nisselson, G. E. Niver, M. L. Norden, Nilan Norris, J. I. Northam, C. O. Oakley, E. G. Olds, P. S. Olmstead, J. G. Osborne, R. F. Passano, Edward Paulson, C. K. Payne, Victor Perlo, J. M. Perotti, L. M. Petit, G. A. D. Preinreich, Harry Press, Elsie Ratkowitz, L. J. Reed, F. V. Reno, J. S. Ripandelli, Selby Robinson, H. G. Romig, A. C. Rosander, Ernest Rubin, H. A. Ruger, P. A. Samuelson, M. M. Sandomire, Max Sasuly, F. E. Satterthwaite, Henry Scheffe, H. L. Schug, H. A. Secrist, Nathan Seiden, W. A. Shelton, R. W. Shephard, W. A. Shewhart, H. M. Shulman, Harry Siller, R. R. Singleton, L. E. Smart, J. H. Smith, G. W. Snedecor, Emma Spaney, Mortimer Spiegelman, Arthur Stein, M. S. Stevens, J. S. Stock, M. M. Torrey, M. N. Torrey, W. R. Van Voorhis, D. F. Votaw, Jr., W. C. Waite, H. M. Walker, W. A. Wallis, A. N. Watson, E. W. Wilson, C. P. Winsor, Jacob Wolfowitz, M. A. Woodbury, W. J. Youden, Joseph Zubin.

The opening session on Saturday afternoon on *The Role of Tests of Significance in Biological Research* was held jointly with the Biometrics Section of the American Statistical Association. Professor E. B. Wilson of the Harvard School of Public Health acted as chairman. The session was in the form of a round table discussion, the principal discussants being: W. Edwards Deming, Bureau of the Census; Harold Hotelling, Columbia University; Lowell J. Reed, Johns Hopkins University; and George W. Snedecor, Iowa State College.

* The list of attendance has been compiled from the registration list supplied by the Director of the New York Convention and Visitors Bureau.

On Saturday evening, under the chairmanship of Dr. Walter A. Shewhart of Bell Telephone Laboratories, a session was held jointly with the Econometric Society on *Theory of Runs and Confidence Intervals*. The following program was presented:

1. *The theory of runs in random data.*
   Harold T. Davis, Northwestern University.
2. *Five time series significance tests based on signs of differences.*
   Geoffrey H. Moore, Rutgers University.
   W. Allen Wallis, Stanford University.
3. *Conference intervals for the unknown median of any type of universe.*
   John H. Smith, University of Chicago.

The morning and afternoon sessions on Sunday on *Numerical Computational Devices* were held jointly with the American Statistical Association, with the co-operation of the Committee on Addresses in Applied Mathematics of the American Mathematical Society. Dr. C. R. Langmuir of the Carnegie Foundation for the Advancement of Teaching acted as chairman of the morning session on *Statistical and Matrix Calculation*. The following papers were presented:

1. *Some matrix methods in least square and other multivariate problems.*
   Harold Hotelling, Columbia University.
2. *The Mallock electrical calculating machine for solving simultaneous linear equations.*
   Elizabeth Monroe Boggs, Cornell University.
3. *Mathematical operations with punched cards.*
   J. C. McPherson, International Business Machines Corporation.
4. *Recent developments in correlation technique.*
   Paul S. Dwyer, University of Michigan.

The subject of the afternoon session was *Mechanical Solution of Differential Equations*. Dr. R. M. Foster of the Bell Telephone Laboratories presided for the following program:

1. *Punch card calculation of orbits.*
   W. J. Eckert, Naval Observatory.
2. *Punch card methods for solving linear differential equations of second order.*
   Martin Schwarzschild, Columbia University.
3. *Differential analyzers.*
   Harold L. Hazen, Massachusetts Institute of Technology.

   Discussants:
   L. S. Dederick, Aberdeen Proving Ground.
   Norbert Wiener, Massachusetts Institute of Technology.

Professor Helen Walker of Columbia University held the chair at the Sunday evening session, a joint session with the American Statistical Association. The following program was given under the title: *On Some Technical Aspects of Sampling.*

1. *On the relative efficiencies of various areal sampling units in population inquiries.*
   M. H. Hansen, Bureau of the Census.
   William Hurwitz, Bureau of the Census.

2. *On the monthly sample survey of unemployment.*
   L. R. Frankel, Work Projects Administration.
   J. S. Stock, Work Projects Administration.
3. *On certain biases in surveys by questionnaire*
   J. Cornfield, Bureau of Labor Statistics
4. *On the relation of probability to sampling.*
   W. G. Madow, Bureau of the Census.
5. *Recent developments in sampling for agricultural statistics*
   G. W. Snedecor, Iowa State College.
   A. J. King, Iowa State College.

Discussants:
   W. G. Cochran, Iowa State College.
   J. A. Greenwood, Duke University.

Another joint session with the American Statistical Association was held on Monday morning. The topic considered was: *What Can the Census Do With Sampling?* Professor L. Edwin Smart of Ohio State University presided for the following program:

1. *An appraisal of the 1940 sampling scheme.*
   T. O. Yntema and Dickson H. Leavens, Cowles Commission for Research in Economics.
2. *Some requirements of sampling design and presentation.*
   W. Edwards Deming, Bureau of the Census
3. *Compromises, losses, and gains brought about by the introduction of sampling.*
   L. E. Truesdell, Bureau of the Census.
4. *The proposed annual sample census.*
   Philip M. Hauser, Bureau of the Census.

Discussants:
   A. N Watson, Curtis Publishing Company.
   F. F. Stephan, Office of Production Management.
   S. A. Stouffer, University of Chicago.

On Monday afternoon, a session was held for the reading of contributed papers on *Probability and Statistics.* Professor Harold Hotelling acted as chairman, and the following papers were read:

1. *Scanning data to determine significance of difference between frequency of an event in contrasted groups.*
   Joseph Zubin, New York State Psychiatric Institute.
2. *Compounding probabilities from independent significance tests.*
   W Allen Wallis, Stanford University.
3. *A class of multivariate distributions.*
   Walter Jacobs, Securities and Exchange Commission.
4. *Definition of the probable error.*
   E. J. Gumbel, New School for Social Research.
5. *A generalized analysis of variance.*
   F. E. Satterthwaite, University of Iowa.
6. *On the power function of the analysis of variance test.*
   Abramham Wald, Columbia University.
7. *Method of computing the roots of cubic and quartic equations by hyperbolic and circular functions.*
   E. E Blanche, Michigan State College.

8 *Additive partition functions.*
J. Wolfowitz, Columbia University
9. *Limited type of probability distribution applied to flood flows* (*Preliminary report*)
B. F. Kimball, New York State Public Service Commission
Abstracts of these papers follow this report.

Professor Harold Hotelling acted as chairman for the session on Tuesday morning, held jointly with the Econometric Society and the American Statistical Association. The program consisted of invited addresses on *Recent Advances in Mathematical Statistics* by Professors Burton H Camp of Wesleyan University and Cecil C. Craig of the University of Michigan.

The session on Tuesday afternoon was held at The Boyce Thompson Institute, Yonkers, New York. It was a joint session with the Biometrics Section of the American Statistical Association on *The Design of Experiments*. Dr. W. J. Youden of The Boyce Thompson Institute acted as chairman and had various experimental designs on display in the greenhouse. Through the courtesy of members of the Institute staff, transportation between the railroad station and the Institute was provided. After the program, tea was served. The following papers were read:

1. *Biological interpretation of interactions.*
W C. Jacobs, Cornell University
2 *Adapting the design to the experiment*
Gertrude M. Cox, North Carolina State College.
3. *Sampling theory when the sampling units are of unequal size*
W. G. Cochran, Iowa State College.
4. *Sampling errors of systematic and random surveys of cover type areas.*
J. G. Osborne, U S. Forest Service.

A luncheon meeting Monday noon was held jointly with the Econometric Society and was attended by ninety-four persons. Professor W. C. Mitchell of Columbia University presided and called on Irving Fisher, Harold Hotelling, W. G. Cochran, and W. A. Wallis for brief remarks

The annual business meeting of the Institute was held late Monday afternoon, with President Hotelling presiding.

The report of the Secretary-Treasurer was read. The report appears on pp. 107–109.

President Hotelling stated that Mr. George W. Petrie, III, had audited the books and records of the Treasurer and found them to be in agreement with the Report presented.

Dr Madow, who acted as teller, reported that the mail balloting had resulted in the election of the following officers for 1942:

President· Professor C. C Craig
Vice-Presidents: Professor A. T Craig
Mr. E. C Molina
Secretary-Treasurer: Professor E G Olds

After discussing various ways of broadening the service of the Institute, a motion was carried which recommended that the Board of Directors appoint committees to study the following matters: junior memberships, local chapters, and advertising for the official journal. Later the Board approved this recommendation and committees were appointed.

EDWIN G. OLDS,
Secretary

REPORT OF THE DALLAS MEETING OF THE INSTITUTE

The twelfth meeting of the Institute was held jointly with the meetings of Section A of the American Association for the Advancement of Science and of the Econometric Society in Dallas on December 29 30, 1941. Professor Dunham Jackson, Secretary of Section A of the A. A. A. S., has kindly sent the following information regarding the meeting:

Sessions of the joint meeting of the Institute of Mathematical Statistics with the Econometric Society and Section A of the A. A. A. S. were held Monday afternoon, December 29, and Tuesday morning and afternoon, December 30, at Southern Methodist University. The number of contributed papers offered on Tuesday was such as to cause extension of the session into the afternoon.

On Monday afternoon addresses were delivered, in accordance with the programs issued in advance, by Professor A. B. Coble of the University of Illinois, retiring Vice President for the Section, on *A Certain Set of Ten Points in Space*, and Professor S. S. Wilks of Princeton University on *Representative Sampling.*

The order of papers on Tuesday was as follows:

1. *On the theory of the tetrahedron.*
   N. A. Court, University of Oklahoma.
2. *A method for integrating the linear hyperbolic equation in three independent variables.*
   E. W. Titt, University of Texas.
3. *On powers of a matrix whose elements are sets of points.*
   S. T. Sanders, Jr., Southwestern Louisiana Institute.
4. *Analytic theory of parametric linear partial differential equations.*
   W. J. Trjitzinsky, University of Illinois.
5. *The theory of the Riesz integral.*
   H. J. Ettlinger, University of Texas.
6. *Obtaining differences from tables which are in the form of punched cards.*
   Harry Pelle Hartkemeier, University of Missouri.
7. *On investment and the valuation of capital.*
   Montgomery D. Anderson, University of Florida.
8. *Advantages of singling out degrees of freedom in analyses of variances.*
   W. D. Baten, Michigan State College.
9. *The incidence of an income tax on saving.*
   Abram Bergson, University of Texas.
10. *Certain tests for randomness applied to data grouped in small sets.*
    Edward L. Dodd, University of Texas.

11. *Stratified sampling* (Preliminary Report).
A. M. Mood, University of Texas.
12. *On convergence factors in convergent integrals.*
Charles N. Moore, University of Cincinnati.
13. *Geometric statement of a fundamental theorem for four-dimensional orthographic axonometry.*
W. H. Roever, Washington University.
14. *A certain non-metric Moore space.*
F. B. Jones, University of Texas.

Abstracts of papers 8, 10, and 11 follow this report.

Papers 1 to 8 inclusive on this list were presented Tuesday morning, and papers 9 to 14 at the afternoon session. In the absence of the authors, papers 10 and 12 were read by title.

The presiding officer Monday afternoon was Professor G. T. Whyburn of the University of Virginia, Chairman of the Section and Vice President of the A. A. A. S. On Tuesday Professor H. J. Ettlinger of the University of Texas presided for papers 1 to 4 inclusive, and Professor S. S. Wilks of Princeton University for the rest of the program.

EDWIN G. OLDS,
Secretary

ANNUAL REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE

On September 2–4, the Institute met at the University of Chicago, in conjunction with meetings of the American Mathematical Society, Mathematical Association of America, and Econometric Society. Sixty-eight members of the Institute attended the meeting.

As mentioned in the 1940 report of the Secretary, the Institute became affiliated with the American Association for the Advancement of Science at the close of 1940. President Hotelling appointed Professor Truman L. Kelley as the representative of the Institute on the Executive Council of the A.A.A.S. for 1941.

On December 29–30, 1941, the Institute held two joint sessions with Section A of the A.A.A.S. and the Econometric Society in connection with the Annual Meeting of the A.A.A.S. at Dallas, Texas. Professor Wilks gave an address at one of the sessions. The report of the Seventh Annual Meeting of the Institute appears on pp. 102–106.

The Institute was invited to send an official representative to the Academic Festival of the University of Chicago, September 27–29, 1941. Mr. John F. Kenney was appointed as the representative of the Institute.

During the past year, the Secretary has received a number of inquiries from members regarding opportunities for doing statistical work in business, government, and industry. While the Institute has no particular organization for such service, the Secretary will be glad to supply information regarding positions which come to his attention.

The Institute has printed an official abstract blank to be used in submitting abstracts for contributed papers. A supply of these blanks can be obtained by writing to the Secretary.

The deaths of two of the members of the Institute have been reported since the last Annual Meeting: Professor James W. Glover, University of Michigan, and Mr. M C. MacLean, Dominion Bureau of Statistics, Ottawa.

The following financial statement covers the period from January 1, 1941 to December 10, 1941:

RECEIPTS

BALANCE, January 1, 1941		$427 47
ROCKEFELLER FOUNDATION GRANT		1,000 00
DUES		2,166 89
SUBSCRIPTIONS		1,158 67
SALES OF BACK NUMBERS		836 29
MISCELLANEOUS		56.93
Total Receipts		$5,646 25

EXPENDITURES

ANNALS OFFICE		
Editorial Expenses 1940	$78.95	
1941	82 02	
Printing	29 75	
		$190.72
WAVERLY PRESS		
Printing and Mailing Annals—4 issues		2,913.09
BACK NUMBERS OFFICE		
Postage and mailing 1940	$27 20	
1941	37.03	
Insurance	33 35	
Purchase of back numbers from H. C. Carver	216 74	
Reprinting 200 copies of Vol V, No. 3	126.09	
		440 41
MEMBERSHIP COMMITTEE		130.58
SECRETARY-TREASURER'S OFFICE		
Filing Case.	$52 06	
Printing and Supplies	150 04	
Postage, telegram, and express	112 78	
Clerical Help	25 40	
		$340.28
PRINTING PROGRAMS FOR MEETINGS		51.74
MISCELLANEOUS		17 89
Total Expenditures		$4,084.71
BALANCE ON HAND, December 10, 1941		1,561 54
		$5,646 25

In comparison with the financial condition of the Institute at the end of 1940, the receipts from dues, subscriptions, and sales of back numbers have increased

nearly two thousand dollars. This is largely due to a net increase of 171 members and 20 subscriptions. Early in the year the Institute received the last thousand dollars of its grant from the Rockefeller Foundation. This source of income has materially assisted the Institute in surviving a period of financial uncertainty. Its loss will be severely felt.

The expenditures of the Institute show a slight decrease, partly due to the fact that fewer back issues of the *Annals* had to be reprinted. An unnecessarily large item of expense is that of the postage which has to be paid because of the slowness of some members and subscribers in paying dues and reporting changes of address. Many copies of the *Annals* have to be reclaimed and mailed a second time. Members could save the Institute considerable expense if they would pay their dues promptly and report change of address well in advance of publication dates of the *Annals*

Financial prospects for 1942 are mixed. The importance of the statistical approach to problems of national defense has caused increased interest in mathematical statistics with the result that many people employed in government service or industry are applying for membership and urging their libraries to subscribe to the *Annals*. On the other hand, delivery to, and collection from, foreign libraries is becoming increasingly difficult, and a marked decrease in the number of foreign subscriptions can be anticipated. Furthermore, operating expenses of the Institute are almost certain to increase as material and labor costs advance. On the whole, it seems very probable that it will require the full co-operation of all the members to avoid operation at a loss during the next calendar year

EDWIN G. OLDS,
Secretary-Treasurer.

December 29, 1941

ABSTRACTS OF PAPERS

I. Presented on December 27, 1941, at the New York Meeting of the Institute

A Generalized Analysis of Variance. FRANKLIN E. SATTERTHWAITE, University of Iowa and Aetna Life Insurance Company.

This paper examines the fundamental principals underlying designs for the analysis of variance. Given several statistics of the type, $\chi_i^2 = \Sigma_i \theta_i^2$, where the θ's are arbitrary orthogonalized linear functions of certain underlying normal data, x_k; a rule is set up for determining a set of m_k as linear functions of the x_k such that $\chi_0^2 = \Sigma(x_k - m_k)^2$ will be independent of the remaining χ_i^2's. Further it is shown that simultaneously with the above, the x's and the θ's may be subjected to certain types of linear restrictions (for the purpose of estimating parameters or otherwise) without disturbing the distributions or the independence relations except for the appropriate reduction in degrees of freedom. The rule used to determine the m's gives results consistent with the standard designs for the analysis of variance. However, it goes further in that one may use weighted rather than simple averages in setting up his design. A practical application of this is the two way analysis of data which are averages and lack homogeniety of variance through constants of proportionality between the variances are known. The two way analysis of incomplete data is another practical problem which is solved by the simple expedient of a zero weight. The use of weighted averages frequently introduces difficulties in estimating parameters, particularly the mean. The combination of the linear restriction concept with standard analysis of variance methods solves this difficulty.

On the Power Function of the Analysis of Variance Test. ABRAHAM WALD, Columbia University.

It is known that the power function of the analysis of variance test depends only on a single parameter, say λ, where λ is a certain function of the parameters involved in the distribution of the sample observations. Let Z be any critical region (subset of the sample space) whose size does not depend on unknown parameters, i.e., it has the same size for all values of the parameters which are compatible with the hypothesis to be tested. It is shown that for any positive c the average power (a certain weighted integral of the power function) of the region Z over the surface $\lambda = c$ cannot exceed the power of the analysis of variance test on the surface $\lambda = c$ (the power of the latter test is constant on the surface $\lambda = c$). P. S Hsu's result, *Biometrika*, January, 1941, pp. 62–68, follows from this as a corollary.

Definition of the Probable Error. E. J. GUMBEL, The New School for Social Research.

The probable error is usually defined either as the semi-interquartile range or as $\frac{2}{3}$ of the standard error. We define it as half of the smallest interval that has the probability $\frac{1}{2}$. For distributions which never increase (decrease), the beginning (end) of this interval is the origin (the median), and the end is the median (the end of the distribution). In general the probable error ρ is the solution of the equations $W(\xi + \rho) - W(\xi - \rho) = \frac{1}{2}$ and $w(\xi + \rho) = w(\xi - \rho)$ where ξ denotes the midpoint of the interval. For symmetrical distributions the first definition remains valid. For the Gaussian distribution the second definition holds besides. The numerical values for the midpoint ξ and the probable error ρ are given for some distributions usual in statistics. The calculation of the standard error of the probable error, which depends upon the distribution $w(x)$, determines whether the probable error is more or less precise than the standard error. For the asymmetrical exponential

distribution the mean and the median have the same precision, and the probable error is more precise than the standard error. For the first law of Laplace, and for Galton's reduced distribution the median and the probable error are more precise than the mean and the standard error For Maxwell's distribution the mean and the probable error are more precise than the median and the standard error.

A Class of Multivariate Distributions. WALTER JACOBS, Security and Exchange Commission, Washington.

The multivariate normal distribution has the property that its probability density is constant along the surface of a hyper-ellipsoid The class of distributions characterized by this property is considered. The form of the characteristic function of any distribution of the class is determined; in this way the parameters of the distribution are shown to be simply related to the first and second moments, when these exist.

Every distribution of the class is the n-variate extension of a univariate symmetrical distribution The method of determining the form of the extension of such a univariate distribution is given. A number of properties of regression for the multivariate normal distribution are shown to hold for any distribution of the class. Among other properties considered is the form of some sampling distributions. Some special cases of interest, including the extensions of the Cauchy distribution and the median law, are discussed briefly.

Methods for Scanning Data to Determine the Significance of the Difference Between the Frequency of an Event in Contrasted Groups. JOSEPH ZUBIN, N. Y. S. Psychiatric Institute, New York.

In many investigations in Psychology, Sociology, Economics and Public Health, there is a need for a quick and ready method for scanning a mass of data in order to select the items that have a significant bearing on the problem under investigation. The statistical procedure for this item analysis consists essentially of evaluating the 2×2 tables which arise when two groups are contrasted for the presence and absence of a given character or event The chi square method or its equivalent, the ratio of the difference between per cents to its standard error, require considerable labor and time and several methods have been proposed for shortening the work. Recently a method was developed which eliminates the need for computing percentages or expected values, the analysis being made with the absolute frequencies. This method depends upon transforming p, the per cent, to the inverse sine function of $\sqrt{p}$. The method is applicable not only to 2×2 tables but can also be made applicable to $2 \times n$ tables and $r \times n$ tables with the aid of simple formulae.

Compounding Probabilities from Independent Significance Tests. W. ALLEN WALLIS, Stanford University.

For combining the probabilities obtained from N independent tests of significance into a single measure, the product of the N independent probabilities provides a criterion which, though rarely ideal, is usually satisfactory. The probability that such a product will be less than Q always exceeds Q, and is the sum of the first N terms in a Poisson series whose parameter is $-\log_e Q$; since this sum is also the probability that a value of χ^2 based on $2N$ degrees of freedom will exceed $-2 \log_e Q$, existing tables of χ^2 may (as R. A. Fisher has pointed out in *Statistical Methods for Research Workers*, section 21.1) be used to test the significance of a product of probabilities. If any of the probabilities have been derived from discontinuous distributions, as is likely with small samples of non-metric data, this method of calculating the probablity of the product fails; in such instances it invariably overstates the probability of the product. Formulas are given for various special cases arising frequently in practice and also for the general case of $D + C$ tests of which D are

based on discontinuous distributions and C on continuous distributions. In several illustrative examples, the overstatement of the joint probability consequent upon neglect of discontinuities is of the order of 100 to 200 per cent.

A Method of Computing the Roots of the General Cubic Equation with Real or Complex Coefficients. ERNEST E. BLANCHE, Michigan State College.

The general cubic equation with real or complex coefficients may readily be reduced to the form $y^3 + 3Hy + G = 0$. Suitable substitutions for y in the reduced equation permit the use of the identities for hyperbolic functions and circular functions: $\sin 3x$, $\cos 3x$, $\sinh 3x$, $\cosh 3x$ and $\sin (u + iv)$. The following classifications may be set up: (A) If $G < 0$ and $H > 0$, only real root is $y = 2\sqrt{H} \sinh z$ where $\sinh 3z = G/2H\sqrt{H} = M$, (B-1) If $G < 0$, $H < 0$, $G/2H\sqrt{-H} \leqq 1$, three real roots, obtained by use of circular identity, $\cos 3x$, (B-2) If $G < 0$, $H < 0$, $G/2H\sqrt{-H} > 1$, only real root is $y = 2\sqrt{-H} \cosh z$ where $\cosh 3z = G/2H\sqrt{-H}$. Complex roots are $-\frac{1}{2} y_1 \pm bi$. The general cubic with complex coefficients has solutions $y_{n+1} = -2\sqrt{H} \sin (u + 2n\pi/3 + iv)$ for $n = 0, 1, 2$, where $\sin (3u + 3iv) = a + bi = M$. For M real, special cases are similar to (A), (B-1) and (B-2).

Limited Type of Probability Distribution Applied to Flood Flows (Preliminary Report). BRADFORD F. KIMBALL, Port Washington, N. Y.

Relative to Gumbel's recent paper on Flood Flows (E. J. Gumbel, "The return period of flood flows," *Annals of Mathematical Statistics*, Vol. 12 (1941)) the author points out that Gumbel's argument that the probability distribution of maximum values does not stem from a limited form of primary probability distribution of the stream flow, is misleading (see page 177, loc. cit.). One might argue for a primary probability distribution of stream flows of the type: $dV = \exp(-\frac{1}{2}u^2)du$ where $u = k[b - \log(a - x)]$, $0 \leqq x \leqq a$, where x is the measure of flow. This increment of x is related to normal probability increment by the linear equation $k\,dx = (a - x)du$. This distribution will not satisfy the condition that von Mises uses in his argument concerning a finite distribution since the cumulative distribution V does not possess a positive derivative of finite order at $x = a$. Also, although x does not have infinite range, the transformed variate u has an infinite range to the right, and will satisfy von Mises' argument for the derivation of the cumulative distribution of the maxima, of the form $\exp[-\exp\{-\alpha(u - u_0)\}]$ in terms of u. The author finds that such a distribution more accurately describes the behavior of maximum annual flood flows than one which ignores the existence of an upper limit a.

Additive Partition Functions. J. WOLFOWITZ, New York City.

Let n_1 and n_2 be positive integers and let

$$m = \max\left(\frac{n_1}{n_1 + n_2}, \frac{n_2}{n_1 + n_2}\right).$$

Let the stochastic variable $V = (v_1, v_2, \cdots v_k)$ be any sequence of positive integers such that $v_1 + v_3 + v_5 + \cdots$ is equal to either one of n_1 and n_2, while $v_2 + v_4 + v_6 + \cdots$ is equal to the other. Two sequences V with the same elements arranged in different order are to be considered distinct and all sequences V are to be assigned the same probability. Such sequences are of statistical importance (Wald and Wolfowitz, *Annals of Math. Stat.*, Vol. 11 (1940). Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions:

1. There exists a pair of positive integers, a and b, such that that

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}$$

2. The series

$$\sum_{i=1}^{\infty} |f(i)| m^{1i}$$

is convergent Then, as n_1 and $n_2 \to \infty$, while n_1/n_2 remains constant, the distribution of the stochastic variable

$$F(V) = \sum_{i=1} f(v_i)$$

approaches the normal distribution. When $f(x) \equiv 1$, $F(V) \equiv U(V)$ (loc. cit., Theorem I). When $f(x) = \log\left(\frac{x^x}{x!}\right)$, $F(V)$ is a statistic introduced by the author (*Amer Math. Soc Bull.* (1941), p. 216).

A similar result holds for partitions of a single integer

II. Presented on December 29, 1941, at the joint session of the Institute, The Econometric Society, and Section A of the A. A. A. S.

Certain Tests for Randomness Applied to Data Grouped into Small Sets. EDWARD L. DODD, University of Texas.

G. Udny Yule, in his paper *A Test of Tippett's Random Sampling Numbers* (*Roy. Stat. Soc. Jour.*, Vol. 101(1938), pp. 167–172), described tests applied to certain sums of the Tippett numbers. Yule regarded the Tippett numbers as not altogether satisfactory.

The tests now to be described, however, involve no summation. For sets of three digits, four classes may be distinguished: The middle number may be the largest, or it may be the least; or the sequence may be monotone increasing or monotone decreasing —here the sequence a, a, a, may be classified with the monotone increasing sequences when $a > 4$; otherwise, with the monotone decreasing sequences Similarly, six consecutive digits in two sets of three digits each give rise to sixteen classes. On the basis of range, sets of two or more of the digits $0, 1, 2 \cdots, 9$ may be separated into ten classes.

Chi-square tests applied by the present author on the basis of the foregoing and similar classifications have not thus far indicated that the Tippett numbers are not satisfactorily random.

Stratified Sampling. A. M. MOOD, University of Texas.

When certain relations between the probabilities $p_1, p_2, \cdots, p_k$ of a multinomial population are known in advance, the technique of stratified sampling provides more efficient estimates of the probabilities than does random sampling. Under certain conditions of stratified sampling, however, the maximum likelihood estimates, n_i/n, of p_i are biased but are unbiased in the limit as the sample size increases. The methods and results of the theory of maximum likelihood require no modification to be made applicable to the problem of estimation in stratified sampling; in fact the results of this theory imply the use of stratified sampling when the conditions for its use obtain.

Advantages of Singling Out Degrees of Freedom in Analyses of Variance. WILLIAM DOWELL BATEN, Michigan Agriculture Experiment Station.

This paper pertains to an experiment involving dummy plots for analyzing effects of placements and fertilizers for cannery peas. Three fertilizers were used at different distances from the pea seeds at planting, the design being a randomized block layout. Advantages are given for breaking up the sum of squares, due to differences between "treatment" means, into sums of squares, each with one degree of freedom. Methods are given for securing the sum of squares involving dummy plots, and obtaining the variances due to main effects and interaction. Interpretations are given for each phase of the analysis.

THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIII, No. 2 — June, 1942

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$5.00 per year. Single copies \$1.50. Back numbers are available at \$5.00 per volume, or \$1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

THE PROGENY OF AN ENTIRE POPULATION[1]

By Alfred J. Lotka

Metropolitan Life Insurance Company

The literature on renewal theory has grown to considerable dimensions, until even admittedly incomplete bibliographies list over 100 titles. But a surprisingly small proportion of these publications exhibits any practical applications to concrete data, and such applications as have been made (e.g by Wicksell, Hadwiger, Rhodes) are for the most part of restricted scope.

Anyone who has been following the development will, I think, feel that this is unfortunate. It has a double disadvantage. On the one hand the purely theoretical discussions emphasize difficulties which in practice may be relatively unimportant, being inherent either in some of the unrealistic *ad hoc* examples discussed, or in the expressions used to fit smooth curves to the basic data, rather than in these data themselves. On the other hand some real difficulties in application to actual data seem to require further clarification.

Several of the applications that have been made, including some of my own, are restricted to following up the "progeny" of a "population element" comprising only individuals all originating at the same time and therefore all of the same age (in the case of industrial equipment installation all made at one point of time). The analysis set forth in the treatment of this special case is competent also to deal with the practically more important case of the progeny of an initial population of given age distribution, though no example of this has hitherto been published.[2] Such an example will now be given, and at the same time this will afford an opportunity to clarify some points in the presentation of the more general case.

Let N_t be the total number of females at time t, and $c_t(a)$ the number comprised within the age limits a and $a + da$. Also, let $m_t(a)$ be the age-specific fertility of females of age a, counting daughters only. If α and ω are, respectively the lower and the upper limit of the female reproductive period, and $B(t)$ the annual births of females, then

$$B(t) = \int_\alpha^\omega N_t c_t(a) m_t(a)\, da. \tag{1}$$

However, it is not in this perfectly general form that the relation is to be applied. The case to be considered is that in which the "initial" population is throughout its "future" development, subject to constant age-specific fertility

[1] Compare A. J. Lotka, "The progeny of a population element," *Am. Jour. Hygiene*, Vol. 8 (1928), p. 875.

[2] An example was given by the writer in an oral communication to the Eighth American Scientific Congress, May 1940, the Proceedings of which have not so far been published

and mortality. If we denote the "initial" time by $t = \omega$ (which we can do since the zero of time is arbitrary), we can then write

$$(2) \qquad B(t) = \int_{\alpha}^{\omega} N_t c_t(a) m_{\omega}(a)\, da, \qquad t > \omega.$$

Also, if $p_{t-a}(a)$ is the probability for a female born at time $\tau = t - a$ of surviving to time t, being then a years old, we have

$$(3) \qquad B(t - a) p_{t-a}(a) = N_t c_t(a),$$

and, in particular, since in the case under consideration $p_{t-a}(a)$ is constant for $t - a > \omega$, i.e., for individuals born after $t = \omega$

$$(4) \qquad B(t - a) p_{\omega}(a) = N_t c_t(a), \qquad t > a + \omega.$$

Now, we have been at liberty for the "future" values of $m_t(a)$ and $p_{t-a}(a)$ to make the arbitrary assumption that they retain their values as of $t = \omega$ and $t - a > \omega$, respectively. But for the "past" of the system under consideration we do not have equal liberty, for any assumption we make must be compatible with

(a) the initial age distribution
(b) equation (1).

We can, however, within these limitations, assume that (4) still holds for $0 < t < \omega$, thus

$$(5) \qquad B(t - a) p_{\omega}(a) = N_t c_t(a), \qquad t > 0.$$

Introducing this in (1) we have

$$(6) \qquad B(t) = \int_{\alpha}^{\omega} B(t - a) p_{\omega}(a) m_t(a)\, da, \qquad t > 0.$$

But we *cannot* now, further assume that

$$(7) \qquad m_t(a) = m_{\omega}(a), \qquad t > 0,$$

for, in general, this would make (6) incompatible with (1).

We can, however, split the integral in (6) into two parts, thus

$$(8) \qquad B(t) = \int_{t}^{\omega} B(t - a) p_{\omega}(a) m_t(a)\, da + \int_{\alpha}^{t} B(t - a) p_{\omega}(a) m_{\omega}(a)\, da,$$

with the assumption, *only in the range* $a < t$,

$$(9) \qquad m_t(a) = m_{\omega}(a), \qquad a < t.$$

Denoting the first integral in (8) by $F(t)$, and contracting $p_{\omega}(a) m_{\omega}(a)$ to $\varphi_{\omega}(a)$, we may write (8) in the form

$$(10) \qquad B(t) = F(t) + \int_{\alpha}^{t} B(t - a) \varphi_{\omega}(a)\, da,$$

$$(11) \qquad = F(t) + \beta(t),$$

with

$$
(12) \qquad \begin{cases} F(t) = 0 & t > \omega \\ F(t) = B(t) & 0 < t < \alpha \end{cases}
$$

and

$$
(13) \qquad B(t) = \int_\alpha^\omega B(t - a)\varphi_\omega(a)\,da, \qquad t > \omega.^{3}
$$

The assumption (9) has a definite physical meaning. The integral in (6) has been so split that the first part, $F(t)$, gives the births of daughters from mothers who themselves were born before $t = 0$, while the second part, $\beta(t)$, gives the births of daughters from mothers born after $t = 0$. Equation (9) therefore expresses the assumption that for mothers born at or after $t = 0$, the age-specific fertilities for ages $a < t$ have the same values $m_\omega(a)$, independent of t, as prevail for $t = \omega$. But at time t there are no mothers of age $a > t$, who were born after $t = 0$. Hence the assumption (9) can be quite simply stated to the effect that the age-specific fertilities $m_\omega(a)$ apply to all mothers born after time $t = 0$. This assumption cannot, in general be made for mothers born before $t = 0$, because it would not, in general, be compatible with the given initial age distribution and at the same time with assumption (5). Hence in the first integral of (8), denoted by $F(t)$ in (10), we must write $m_t(a)$, not $m_\omega(a)$.

Equation (10) is of the form discussed by G. Herglotz,[4] who writes its solution, for $t > 0$, in the form of an exponential series.

$$
(14) \qquad B(t) = \Sigma Q_j e^{r_j t}
$$

where the exponents r_j are the roots of the characteristic equation,

$$
(15) \qquad \Phi(r) = \int_\alpha^\omega e^{-ra}\varphi_\omega(a)\,da = 1,
$$

while the coefficients Q_j are given by

$$
(16) \qquad Q_j = \frac{\int_0^\omega F(t)e^{-r_j t}\,dt}{\int_\alpha^\omega a e^{-ra}\varphi_\omega(a)\,da}.
$$

There is only one real root of (14), since $\varphi_\omega(a) \geqq 0$, for all values of a. For complex roots it is convenient to write the corresponding terms of the series (14) in trigonometric form

$$
(17) \qquad Qe^{rt} = 2Ue^{ut}\cos vt - 2Ve^{ut}\sin vt,
$$

$$
(18) \qquad = 2\sqrt{(U^2 + V^2)}e^{ut}\cos(vt + \theta),
$$

---

[3] Since $\varphi_\omega(a) = 0$ for $a > \omega$.

[4] *Math. Annalen*, Vol 65 (1908), pp 87 et seq.

where

$$
\begin{aligned}
&\tan\theta = V/U, \\
(19) \qquad &\cos\theta = \frac{U}{\sqrt{U^2 + V^2}}, \\
&\sin\theta = \frac{V}{\sqrt{U^2 + V^2}},
\end{aligned}
$$

and

$$
(20) \qquad U = \frac{RG + SH}{G^2 + H^2},
$$

$$
(21) \qquad V = \frac{RH - SG}{G^2 + H^2},
$$

in which

$$
(22) \qquad G = \int_\alpha^\omega a e^{-ua} \cos va \, \varphi_\omega(a) \, da,
$$

$$
(23) \qquad H = \int_\alpha^\omega a e^{-ua} \sin va \, \varphi_\omega(a) \, da,
$$

$$
(24) \qquad R = \int_0^\omega e^{-ut} \cos vt \, F(t) \, dt,
$$

$$
(25) \qquad S = \int_0^\omega e^{-ut} \sin vt \, F(t) \, dt.
$$

For purposes of numerical application to the problem here considered, we must express the annual births $B(t)$ for $t < \omega$ in terms of the given "initial" age distribution at time ω.

We have, generally

$$
(26) \qquad B(t - a) = \frac{N_t c_t(a)}{p_{t-a}(a)} = \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)},
$$

since individuals of age a at time t, are $a + \omega - t$ years old at time ω.

Introducing the relation (26) in (10) we have

$$
(27) \qquad B(t) = F(t) + \int_\alpha^t \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)} p_\omega(a) m_\omega(a) \, da,
$$

and

$$
(28) \qquad F(t) = B(t) - \int_\alpha^t \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)} \varphi_\omega(a) \, da,
$$

$$
(29) \qquad = \frac{N_\omega c_\omega(\omega - t)}{p_\omega(\omega - t)} - \int_\alpha^t \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)} \varphi_\omega(a) \, da,
$$

$$
(11a) \qquad = B(t) - \beta(t)
$$

Note that, in computing the integral $\beta(t)$ for any particular value of t, the argument of the function c_ω runs from $\alpha + \omega - t$ to ω. Thus, for example, if the zero of time is 1865 and $t = \omega$ is at 1920, then, in computing $F(35)$, i e, the value of F for 1900, the range of the argument of c_ω in the integral will be from $10 + 55 - 35$ to 55, i.e., from 30 to 55

Numerical Example. By way of a numerical illustration these principles will now be applied to a concrete case. We shall start with the age distribution of the white female population of the United States as constituted in 1920, for which previous publications furnish some of the required data, including the real root and the first three pairs of complex roots of the characteristic equation.

From this "initial" age distribution in 1920 it is necessary first of all to compute the auxiliary function $F(t)$ for the 55 years prior to 1920. The first term $B(t)$ in the right hand member of (28) is very easily computed for successive values of t from the relation (5a), which simply expresses the fact that persons a years old in the year ω, i.e., 1920, are the survivors of the $B(\omega - a)$ persons born in the year $\omega - a$.

$$N_\omega c_\omega(a) = B(\omega - a)p_\omega(a). \tag{5a}$$

In the diagram Fig. 1, which is drawn in stereographic projection, the age distribution of the (white female) population of the United States in 1920 is represented as plotted in a plane reaching forward at right angles to the plane of the paper. Successive points of $B(t)$ for $0 \leq t \leq \omega$, have been computed "by survivals" according to (5a) and plotted as a curve in the plane of the paper "at the back" of the diagram The arrows indicate for a selected point, namely age 25 in 1920, the path of the computation according to equation (5a.)

The second term $\beta(t)$ in the expression (11a) for $F(t)$ was computed from the age distribution in 1920, the rates of survival from previous years into 1920,[5] and the age-specific fertility at each age in the reproductive period, 10 to 55, on the basis of the relation (28). The results, for this second term in the expression for $F(t)$ computed for every fifth calendar year back of 1920 to 1875 and interpolated for intervening years,[6] were also plotted as a curve in the rear plane of the diagram. The shaded area in the curve for the age distribution in 1920, and the arrows leading from this shaded area to the curve

$$\beta(t) = \int_\alpha^t B(t - a)\varphi_\omega(a)\, da \tag{10, 11}$$

$$= \int_\alpha^t \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)} \varphi_\omega(a)\, da, \tag{29, 11a}$$

indicate in this case the path of the computation according to equation (28).

[5] Using the Foudray life table for white females in 1919–1920 In the first quinquennial age group, the following values were used,

$$p(0.5) = .9460 \qquad p(2.5) = .9135$$

$$p(1\ 5) = 9235. \qquad p(3.5) = .9080 \qquad p(4\ 5) = .9040$$

[6] This term vanishes for $t < 10$, i.e., back of 1875.

From these two curves, taking differences, the curve of $F(t) = B(t) - \beta(t)$ was plotted, as shown.

With the values of $F(t)$ thus obtained, we may proceed, by formulae (14) to (25), to compute values of $B(t)$ for all values of $t > 0$. So far as the period 1865 to 1920, corresponding to $0 < t < \omega$, is concerned, this merely means that we have an analytical expression to fit what is essentially a fundamental datum of the problem. For values of $t > \omega$ the formula gives us a continuation of the function $B(t)$ for all future time so long as the given age-specific fertility and mortality holds.

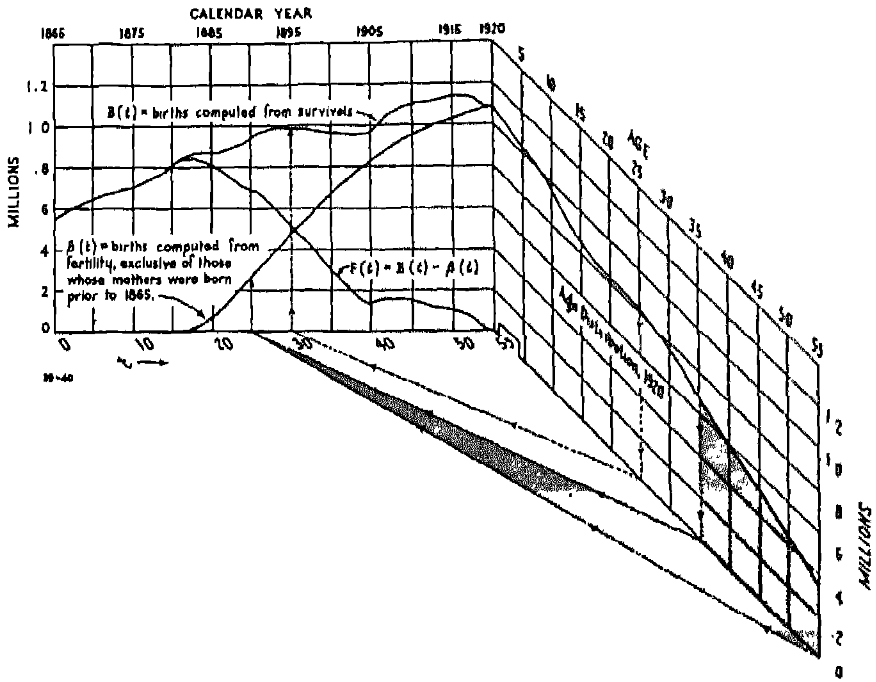


FIG. 1. Graph illustrating computation of auxiliary function $F(t)$ from "initial" age distribution.

The final results of this computation are exhibited in Figs. 2, 3 and 4. Of these, Fig. 2 exhibits the first, second and third oscillatory components for the period from 1890 forward. It will be seen that the waves are heavily damped, so that after a relatively short period the aperiodic component dominates the course of events.

Fig. 3 exhibits, for the years from 1865 to 1920, i.e., for the period $0 < t < \omega$, the aperiodic component (in a dashed line) and, as indicated by small circles, the sum of this component plus the three oscillatory components. It will be seen that from about 1890 forward the points so obtained follow rather closely the value $B(t)$ derived by survivals from the age distribution in 1920.

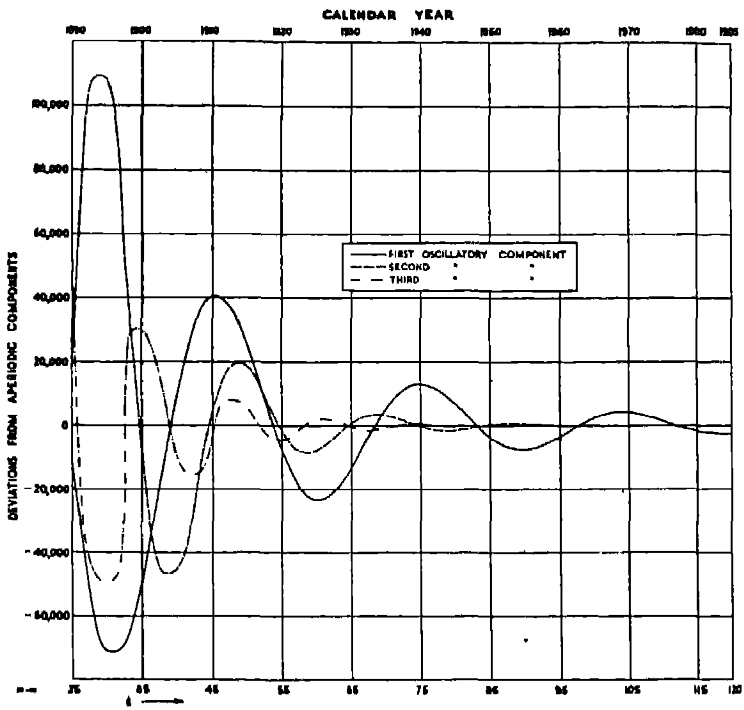


FIG. 2 First three oscillatory components of total annual births

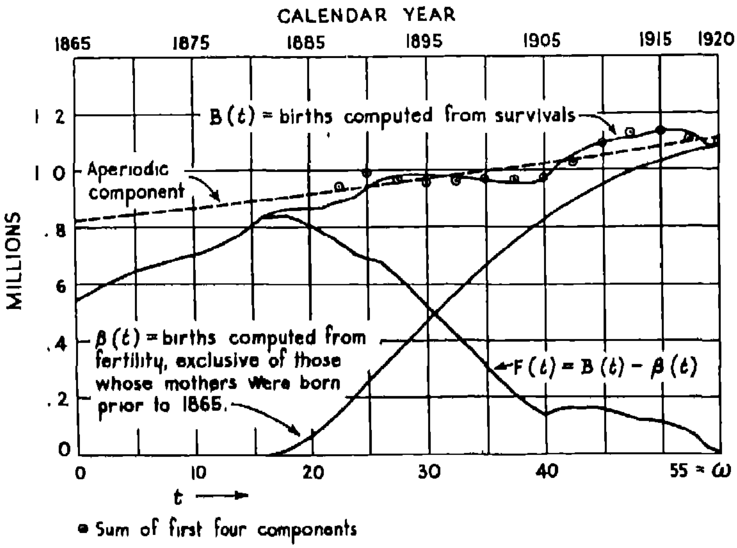


FIG. 3 Graph of functions $B(t)$, $\beta(t)$, and $F(t)$ for $0 < t < \omega$, i.e., for 1865 to 1920, together with aperiodic component; also, summation of aperiodic and first three oscillatory components.

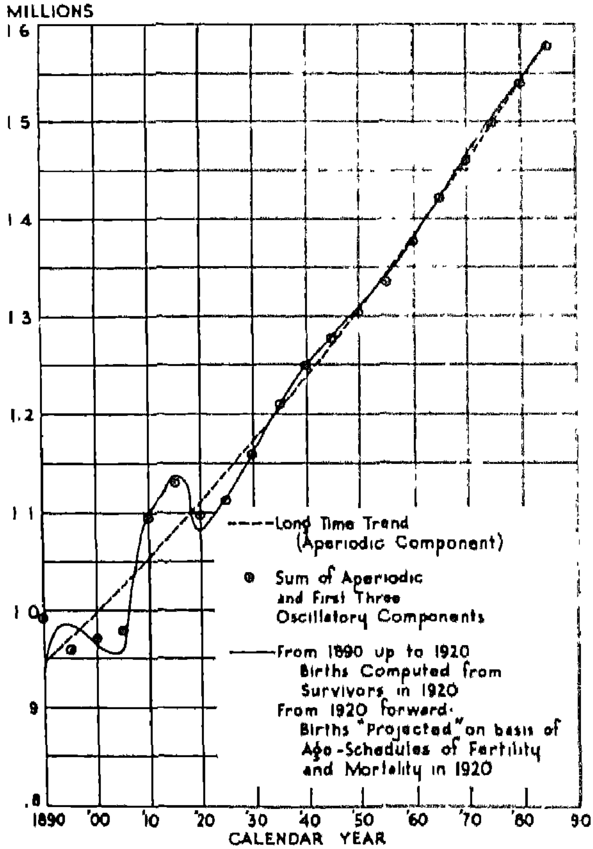


FIG. 4. Sum of aperiodic and three oscillatory terms of series solution compared with results of "step by step" computation of annual births.

TABLE I

Constants of the Series Solution (14) of Integral Equation (10) to Third Oscillatory Component Inclusive $t = 0$ at 1865

Function	Aperiodic Component	Oscillatory Components		
		First	Second	Third
u	$.543 \times 10^{-2}$	$-.386 \times 10^{-1}$	-8.731×10^{-2}	-9.804×10^{-2}
v	0	21.448×10^{-2}	31.542×10^{-2}	48.849×10^{-2}
G	28.226	25.768	51.225	37.008
H	0	14.938	-18.637	17.266
R	23.262×10^{6}	-17.863×10^{5}	-37.196×10^{6}	11.684×10^{6}
S	0	-31.508×10^{5}	16.827×10^{6}	-16.543×10^{6}
U	82.416×10^{4}	-10.494×10^{4}	-74.679×10^{4}	88.014×10^{3}
V	0	61.442×10^{3}	-56.787×10^{3}	48.808×10^{4}

Prior to about 1890, four components alone are quite inadequate, and the corresponding points have been omitted from the diagram The lack of concordance, with such limited components, is inconsequential in this part of the series, since the purpose of this part of the work was merely to compute the auxiliary function $F(t)$, and the fit obtained for $B(t)$ in this range, so far as it goes, is merely a by-product, the main interest being in the course of $B(t)$ for $t > \omega$, i.e., in the years following 1920

This course is charted in Fig 4, in which the points obtained by the series solution (14) of (10) are again shown as small circles, while the fully drawn curve is derived from my previous publication "The Progressive Adjustment of Age Distribution to Fecundity."[7] The annual births in that case were obtained "step by step" by computing age distributions by survivals for successive

TABLE II

United States White Female Population 1920, Observed, Also, the Same Projected Forward for Later Years[8]

Year	Population, thousands	Births, thousands	Birth rate per 1,000 per annum
1920	49,390	1,082	23 32
1930	51,727	1,162	22.46
1940	56,910	1,252	22.00
1950	61,639	1,307	21.20
1960	65,835	1,379	20.95
1970	69,829	1,465	20 98
1975	71,828	1,504	20 94
1980	73,850	1,543	20.89
1985	75,902	1,584	20.87

quinquennial periods, and applying to the reproductive age groups, in each case, the values of the reproductivity $m_\omega(a)$.

It will be seen that the points obtained by the solution (14) follow very closely those computed "step by step," although in the computation of the latter an approximation was made, using pivotal values of $p_\omega(a)$ for the several quinquennial age groups. A slight error introduced in this way would tend to be cumulative, and perhaps accounts for the fact that towards the end of the period covered (1985), the two sets of values diverge slightly. Even so, in 1985, the divergence is only about .4 percent.

The series solution has, of course, the advantage that it gives directly the result for any particular point of time, whereas the "step by step" method re-

[7] *Jour. Washington Acad Sci.*, Vol. 16 (1926), p. 505.

[8] Calculated step by step from survival ratios and age specific fertilities, both held constant as of 1920 (reproduced for ready reference from *Jour. Wash. Acad. Sci.*, Vol. 16, p. 505).

quires the computation of the annual births for all intervening points in order to obtain the result for the chosen point of time.

Furthermore, the series tells us at once that the course of events is of the nature of a trend proceeding in geometric progression upon which are superposed a series of damped oscillations, of which the fundamental has a wave length equal approximately to the mean length of one generation from mother to daughter, i.e., about 28 years.

Alternative procedure. The procedure set forth in the preceding sections involves not only arbitrary assumptions regarding the values of $p(a)$ and $m(a)$ for "future" time, which are fundamental to the problem under consideration, but involves further incidental assumptions regarding their values prior to the "initial" condition at the instant denoted by $t = \omega$. These incidental assumptions are in a sense superfluous, since the future history of the system is completely determined by the initial age distribution and the assumed "future" values of $p(a)$ and $m(a)$. The additional assumptions were introduced merely for the purpose of translating the initial age distribution into a series of values of $B(t)$ for $0 < t < \omega$, i.e., prior to the given initial age distribution.

In actual fact the age distribution at time $t = \omega$ did not arise in the manner assumed; actually both $p(a)$ and $m(a)$ undoubtedly varied in the period 1865 to 1920, and migration also affected the situation. The quantity $F(t)$ introduced in equation (10) is, in fact, a purely auxiliary function having no direct relation to the biological events at time $t < \omega$.

An alternative procedure which would avoid these conflicts, and introduce assumptions only regarding "future" values of $p(a)$ and $m(a)$, would be to compute $B(t)$ step by step over the period from $B(1920)$ to $B(1920 + \omega) = B(1975)$.

Placing the zero of time $t = 1920$ this would give $B(t)$ for $0 < t < \omega$. For $t > \omega$ we should have, simply

$$B(t) = \int_{\alpha}^{\omega} B(t - a)\varphi_{1920}(a)\, da, \qquad t > \omega,$$

using, in the evaluation of the integral, the values of $B(t - a)$ obtained by the step by step process.

We could here also split the integral into two parts

$$\begin{aligned} B(t) &= \int_{t}^{\omega} B(t - a)\varphi_{1920}(a)\, da + \int_{\alpha}^{t} B(t - a)\varphi_{1920}(a)\, da \\ &= F(t) + \int_{\alpha}^{t} B(t - a)\varphi_{1920}(a)\, da. \end{aligned}$$

But the function $\varphi_{1920}(a)$ is now the same in the two integrals, and there is no occasion, in this case, for distinguishing the two parts of the integral.

If this procedure is adopted, its application to the course of $B(t)$ for $t > \omega$,

i.e. beyond 1975, is of minor interest, for by that time it has practically settled down to the aperiodic (exponential) component, the oscillations being greatly damped down. The major interest in the result of a computation carried out by this procedure would be in the fitting of a series of the form (14) to the function $B(t)$ in the range 1920 to 1975, which, in this setting, figures as a known "arbitrary" function.

Of the two alternative procedures the one carried out in detail in the text and the numerical example is of greater interest, as exhibiting in greater generality the application of the Hertz-Herglotz solution.

BIBLIOGRAPHY

A few more titles may be added to previously published bibliographies (see these Annals, Vol. 10 (1939), p. 22; Vol. 12 (1941), p 266).

1935 W. Möschler, "Untersuchungen über Eintrittsgewinn und Fehlbetrag einer Versicherungskasse," *Bull. de l'Association des Actuaires Suisses*, October (1935), p. 129

1936 Ali Afzalipour, *Contribution à l'étude de la théorie mathématique de la démographie*, Thèse, 1936

1938 H. Hadwiger, "Ein Konvergenzkriterium für Erneuerungszahlen," *Skandinavisk Aktuarietidskrift*, Vol. 3-4, p. 226.

H. Hess, *Anwendung der logistischen Funktion in der mathematischen Bevölkerungstheorie*, Inaugural Dissertation der philosophischen Fakultät der Universität Bern.

W. M. Dawson and W Edwards Deming, "On the problem of natural increase," *Growth*, Vol. 2, p. 319.

1940 A. W. Brown, "A note on the use of a Pearson Type III function in renewal theory," *Annals of Math Stat.*, Vol. 11 (1940), p. 448.

1940 E. Zwinggi, "Entwicklung von Personengesamtheiten—Zusammenfassender Bericht," *Twelfth International Congress of Actuaries*, Vol. 3, p. 263

1940 E. Keinanen, "Uber die Altersverteilung der Bevölkerung," *Twelfth International Congress of Actuaries*, Vol. 3, p. 305.

1940 S. S. Townsend, "Some observations on the internal variation in groups of lives assured in industrial assurance and in the general population of England and Wales," *Twelfth International Congress of Actuaries*, Vol 3, p. 319.

1940 R Tarjan, "Untersuchungen über den Kapitalbedarf des Lebensversicherungsgeschäftes," *Twelfth International Congress of Actuaries*, Vol. 3, p 335.

1940 M. Presburger, "Sur l'étude générale des collectivités de personnes," *Twelfth International Congress of Actuaries*, Vol. 3, p. 353.

1940 A Maret, "Direkte Berechnung der Vorgangsfunktionen einer offenen Gesamtheit," *Twelfth International Congress of Actuaries*, Vol. 3, p. 387.

1940 E. Zwinggi, "Über Zusammenhänge zwischen der technischen Stabilität einer Sozialversicherungskasse und der Entwicklungsformel für den Versichertenbestand," *Twelfth International Congress of Actuaries*, Vol. 3, p. 395.

1940 H. Hadwiger and W. Wegmüller, "Entwicklung und Umsichichtung von Personengesamtheiten," *Twelfth International Congress of Actuaries*, p. 369.

1940 W. Dobbernack and G. Tietz, "Die Entwicklung von Personengesamtheiten vom Standpunkt der Sozialersicherungstechnik," "Zwölfter Internationaler Kongress Der Versicherungs-Mathematiker," Vol. 4, p. 233.

1941 R C. Geary, "Irish population prospects considered from the viewpoint of reproduction rates," *Statistical and Social Inquiry Society of Ireland*, 1941.

1941 H. HADWIGER, "Eine Formel der mathematischen Bevolkerungstheorie," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Vol. 41, p. 67

L. FERAUD, "Le renouvellement, quelques problèmes connexes et les equations intégrales de cycle fermé," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Vol. 41, p. 81

W. FELLER, "On the integral equation of the renewal theory," *Annals of Math. Stat.*, Vol. 12 (1941), p. 243.

HARRO BERNARDELLI, "Population waves," *Journal of the Burma Research Society*, Vol. 31.

ASYMPTOTICALLY SHORTEST CONFIDENCE INTERVALS[1]

By Abraham Wald[2]

Columbia University

The theory of confidence intervals, based on the classical theory of probability, has been treated by J. Neyman.[3] While Neyman considers the case of small samples, we shall deal here with the limit properties of the confidence intervals if the number of observations approaches infinity.

1. Definitions. We will start with some of Neyman's definitions. Let $f(x, \theta)$ be the probability density function of a variate x involving an unknown parameter θ. Denote by E_n a point of the n-dimensional sample space of n independent observations on x. If $\rho(E_n)$ denotes for each E_n a subset of the real axis, the symbol $P[\rho(E_n)c\theta' \mid \theta'']$ will denote the probability that $\rho(E_n)$ contains θ' under the hypothesis that θ'' is the true value of the parameter. Let $\underline{\theta}(E_n)$ and $\bar{\theta}(E_n)$ be two real functions defined over the whole sample space such that $\underline{\theta}(E_n) \leq \bar{\theta}(E_n)$. The interval $\delta(E_n) = [\underline{\theta}(E_n, \bar{\theta}(E_n)]$ is called a confidence interval of θ corresponding to the confidence coefficient α $(0 < \alpha < 1)$ if $P[\delta(E_n)c\theta \mid \theta] = \alpha$ for all values of θ.

The interval function $\delta(E_n)$ is called a shortest confidence interval of θ corresponding to the confidence coefficient α if

(a) $P[\delta(E_n)c\theta \mid \theta] = \alpha$ for all values of θ, and

(b) for any interval function $\delta'(E_n)$ which satisfies the condition (a) we have

$$P[\delta(E_n)c\theta' \mid \theta''] \leq P[\delta'(E_n)c\theta' \mid \theta''],$$

for arbitrary values θ' and θ''.

The interval function $\delta(E_n)$ is called a shortest unbiased confidence interval of θ if the following three conditions are fulfilled:

(a) $P[\delta(E_n)c\theta \mid \theta] = \alpha$ for all values of θ.

(b) $P[\delta(E_n)c\theta' \mid \theta''] \leq \alpha$ for all values of θ' and θ''.

(c) For any interval function $\delta'(E_n)$ for which the conditions (a) and (b) are satisfied, we have

$$P[\delta(E_n)c\theta' \mid \theta''] \leq P[\delta'(E_n)c\theta' \mid \theta''],$$

for all values of θ' and θ''.

For any relation R we shall denote by $P(R \mid \theta)$ the probability that R holds under the hypothesis that θ is the true value of the parameter. Similarly for

---

[1] Presented at a joint meeting of the Institute of Mathematical Statistics and the American Mathematical Society in Hanover, September, 1940.

[2] Research under a grant-in-aid from the Carnegie Corporation of New York.

[3] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy Soc. London*, Vol. 236 (1937), pp. 333–380.

any region Q_n of the n-dimensional sample space the symbol $P(Q_n \mid \theta)$ will denote the probability that the sample point E_n falls in Q_n under the hypothesis that θ is the true value of the parameter.

In all that follows we shall denote a region of the n-dimensional sample space by a capital letter with the subscript n.

A real function $\bar{\theta}(E_n)$ is called a best upper estimate of θ if the following two conditions are fulfilled:

(a) $P[\theta \leq \bar{\theta}(E_n) \mid \theta] = \alpha$ for all values of θ.

(b) For any function $\bar{\theta}'(E_n)$ which satisfies the condition (a) we have

$$P[\theta' \leq \bar{\theta}(E_n) \mid \theta''] \leq P[\theta' \leq \bar{\theta}'(E_n) \mid \theta'']$$

for all values θ' and θ'' for which $\theta' \geq \theta''$.

A real function $\underline{\theta}(E_n)$ is called a best lower estimate of θ if the following two conditions are fulfilled:

(a) $P[\theta \geq \underline{\theta}(E_n) \mid \theta] = \alpha$ for all values of θ.

(b) For any function $\underline{\theta}'(E_n)$ which satisfies the condition (a) we have

$$P[\theta' \geq \underline{\theta}(E_n) \mid \theta''] \leq P[\theta' \geq \underline{\theta}'(E_n) \mid \theta'']$$

for all values of θ' and θ'' for which $\theta' \leq \theta''$.

We will extend the above definitions of Neyman to the limit case when n approaches infinity.

DEFINITION I: *A sequence of interval functions* $\{\delta_n(E_n)\}$ $(n = 1, 2, \cdots)$ *is called an asymptotically shortest confidence interval of* θ *if the following two conditions are fulfilled*:

(a) $P[\delta_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(b) *For any sequence of interval functions* $\{\delta_n'(E_n)\}$ $(n = 1, 2, \cdots, ad\ inf.)$ *which satisfies* (a), *the least upper bound of*

$$P[\delta_n(E_n) c \theta' \mid \theta''] - P[\delta_n'(E_n) c \theta' \mid \theta'']$$

with respect to θ' *and* θ'' *converges to zero as* $n \to \infty$.

DEFINITION II: *A sequence of interval functions* $\{\delta_n(E_n)\}$ *is called an asymptotically shortest unbiased confidence interval of* θ *if the following three conditions are fulfilled*:

(a) $P[\delta_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(b) *The least upper bound of* $P[\delta_n(E_n) c \theta' \mid \theta'']$ *with respect to* θ' *and* θ'' *converges to* α *with* $n \to \infty$.

(c) *For any sequence of interval functions* $\{\delta_n'(E_n)\}$ *which satisfies the conditions* (a) *and* (b), *the least upper bound of*

$$P[\delta_n(E_n) c \theta' \mid \theta''] - P[\delta_n'(E_n) c \theta' \mid \theta'']$$

with respect to θ' *and* θ'' *converges to zero with* $n \to \infty$.

DEFINITION III: *A sequence of real functions* $\{\bar{\theta}_n(E_n)\}$ $(n = 1, 2, \cdots, ad\ inf.)$ *is called an asymptotically best upper estimate of* θ *if the following two conditions are fulfilled*:

(a) $P[\theta \leq \bar{\theta}_n(E_n) \mid \theta] = \alpha$ *for all values of* θ.

(b) *For any sequence of functions* $\{\bar{\theta}'_n(E_n)\}$ *which satisfies* (a) *the least upper bound of*

$$P[\theta' \leq \bar{\theta}_n(E_n) \mid \theta''] - P[\theta' \leq \bar{\theta}'_n(E_n) \mid \theta'']$$

in the domain $\theta' \geq \theta''$ *converges to zero with* $n \to \infty$.

DEFINITION IV: *A sequence of real functions* $\{\underline{\theta}_n(E_n)\}$ *is called an asymptotically best lower estimate of* θ *if the following two conditions are fulfilled:*

(a) $P[\theta \geq \underline{\theta}_n(E_n) \mid \theta] = \alpha$ *for all values of* θ.

(b) *For any sequence of functions* $\{\underline{\theta}'_n(E_n)\}$ *which satisfies* (a) *the least upper bound of*

$$P[\theta' \geq \underline{\theta}_n(E_n) \mid \theta''] - P[\theta' \geq \underline{\theta}'_n(E_n) \mid \theta'']$$

in the domain $\theta' \leq \theta''$ *converges to zero with* $n \to \infty$.

2. Two Propositions. PROPOSITION I: *Let* $\{W_n(\theta)\}$ $(n = 1, 2, \cdots, ad\ inf.)$ *be for each* θ *a sequence of regions such that the following two conditions are fulfilled:*

(a) $P[W_n(\theta) \mid \theta] = 1 - \alpha$ *for all values of* θ.

(b) *For any sequence of regions* $\{Z_n(\theta)\}$ *which satisfies* (a) *the least upper bound of*

$$P[Z_n(\theta') \mid \theta''] - P[W_n(\theta') \mid \theta'']$$

in the domain $\theta' \geq \theta''(\theta' \leq \theta'')$ *converges to zero with* $n \to \infty$.

Denote by $\rho_n(E_n)$ *the set of all values of* θ *for which* E_n *does not lie in* $W_n(\theta)$. *Then we have*

(c) $P[\rho_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(d) *For any sequence of set functions* $\{\rho'_n(E_n)\}$ *which satisfies* (c), *the least upper bound of*

$$P[\rho_n(E_n) c \theta' \mid \theta''] - P[\rho'_n(E_n) c \theta' \mid \theta'']$$

in the domain $\theta' \geq \theta''(\theta' \leq \theta'')$ *converges to zero with* $n \to \infty$.

PROPOSITION II: *Let* $\{W_n(\theta)\}$ *be for each* θ *a sequence of regions such that the following three conditions are fulfilled:*

(a) $P(W_n(\theta) \mid \theta] = 1 - \alpha$ *for all values of* θ.

(b) *The greatest lower bound of* $P[W_n(\theta') \mid \theta'']$ *converges to* $1 - \alpha$ *with* $n \to \infty$.

(c) *For any sequence* $\{W'_n(\theta)\}$ *which satisfies* (a) *and* (b), *the least upper bound of*

$$P[W'_n(\theta') \mid \theta''] - P[W_n(\theta') \mid \theta'']$$

with respect to θ' *and* θ'' *converges to* 0 *with* $n \to \infty$.

Denote by $\rho_n(E_n)$ *the set of all values of* θ *for which* E_n *does not lie in* $W_n(\theta)$. *Then we have*

(d) $P[\rho_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(e) *The least upper bound of* $P[\rho_n(E_n) c \theta' \mid \theta'']$ *converges to* α *with* $n \to \infty$.

(f) *For any sequence of setfunctions* $\{\rho'_n(E_n)\}$ *which satisfies* (d) *and* (e), *the least upper bound of*

$$P[\rho_n(E_n) c \theta' \mid \theta''] - P[\rho'_n(E_n) c \theta' \mid \theta'']$$

with respect to θ' *and* θ'' *converges to* 0 *with* $n \to \infty$.

The validity of the above propositions follows easily from the identity

$$P[\rho_n(E_n)c\theta' \mid \theta''] = 1 - P[W_n(\theta') \mid \theta''].$$

3. Assumptions on the probability density function. For any function $\psi(x)$ denote by $E_\theta\psi(x)$ the expected value of $\psi(x)$ under the assumption that θ is the true value of the parameter, i.e.

$$E_\theta\psi(x) = \int_{-\infty}^{+\infty} \psi(x)f(x, \theta)\, dx.$$

For any x, for any positive δ, and for any real value θ' denote by $\varphi_1(x, \theta', \delta)$ the greatest lower bound, and by $\varphi_2(x, \theta', \delta)$ the least upper bound of $\frac{\partial^2}{\partial\theta^2} \log f(x, \theta)$ in the interval $\theta' - \delta \leq \theta \leq \theta' + \delta$.

Throughout this paper the following assumptions on $f(x, \theta)$ will be made:

Assumption I: *The expectation* $E_{\theta'} \frac{\partial}{\partial\theta} \log f(x, \theta'')$ *is a continuous function of* θ' *and* θ'', *and for any pair of sequences* $\{\theta'_n\}$ *and* $\{\theta''_n\}$ $(n = 1, 2, \cdots, ad\ inf.)$ *for which*

$$\lim_{n=\infty} E_{\theta'_n} \frac{\partial}{\partial\theta} \log f(x, \theta''_n) = 0$$

also

$$\lim_{n=\infty} (\theta'_n - \theta''_n) = 0.$$

Furthermore

$$E_{\theta'}\left[\frac{\partial}{\partial\theta} \log f(x, \theta'')\right]^2$$

is a bounded function of θ' *and* θ'', *and* $E_\theta\left[\frac{\partial}{\partial\theta} \log f(x, \theta)\right]^2 = d(\theta)$ *has a positive lower bound.*

Assumption II: *There exists a positive value* k_0 *such that the expectations* $E_{\theta'}\varphi_1(x, \theta'', \delta)$ *and* $E_{\theta'}\varphi_2(x, \theta'', \delta)$ *are uniformly continuous functions of* θ', θ'' *and* δ *where* δ *takes only values for which* $|\delta| \leq k_0$. *Furthermore it is assumed that* $E_{\theta'}[\varphi_i(x, \theta'', \delta)]^2$ $(i = 1, 2)$ *are bounded functions of* θ', θ'' *and* δ $(|\delta| \leq k_0)$.

Assumption III: *The relations*

$$\int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(x, \theta)\, dx = \int_{-\infty}^{+\infty} \frac{\partial^2}{\partial\theta^2} f(x, \theta)\, dx = 0$$

hold.

The above assumption means simply that we may differentiate with respect to θ under the integral sign In fact

$$\int_{-\infty}^{+\infty} f(x, \theta)\, dx = 1$$

identically in θ. Hence

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} f(x, \theta)\, dx = \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{+\infty} f(x, \theta)\, dx = 0.$$

Differentiating under the integral sign, we obtain the relations in Assumption III.

ASSUMPTION IV: *There exists a positive* η *such that*

$$E_\theta\left[\frac{\partial}{\partial\theta}\log f(x, \theta)\right]^{2+\eta}$$

is a bounded function of θ.

4. Some theorems. The assumptions on $f(x, \theta)$ made in this paper become identical with the assumptions I–IV formulated in a previous paper[4] if a certain set ω involved in those assumptions is put equal to the whole real axis $(-\infty, +\infty)$ Hence we can make use of all results obtained in that paper putting $\omega = (-\infty, +\infty)$. Among others, the following statements have been proved there.

(A) Denote $\sum_{\alpha=1}^{n} \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta}\log f(x_\alpha, \theta)$ by $y_n(\theta, E_n)$ and let $R_n(\theta)$ be the region defined by the inequality $y_n(\theta, E_n) \geq A_n(\theta)$ where $A_n(\theta)$ is chosen such that $P[R_n(\theta) \mid \theta] = 1 - \alpha$. Then for any sequence of regions $\{Z_n(\theta)\}$ for which $P[Z_n(\theta) \mid \theta] = 1 - \alpha$, the least upper bound of

$$P[Z_n(\theta') \mid \theta''] - P[R_n(\theta') \mid \theta'']$$

in the set $\theta'' \geq \theta'$ converges to 0 with $n \to \infty$.

(B) Let $S_n(\theta)$ be the region defined by the inequality $y_n(\theta, E_n) \leq B_n(\theta)$ where $B_n(\theta)$ is defined such that $P[S_n(\theta) \mid \theta] = 1 - \alpha$. Then for any sequence of regions $\{Z_n(\theta)\}$ for which $P[Z_n(\theta) \mid \theta] = 1 - \alpha$, the least upper bound of

$$P[Z_n(\theta') \mid \theta''] - P[S_n(\theta') \mid \theta'']$$

in the set $\theta'' \leq \theta'$ converges to 0 with $n \to \infty$.

(C) Denote by $T_n(\theta)$ the region defined by $\mid y_n(\theta, E_n) \mid \geq C_n(\theta)$ where $C_n(\theta)$ is chosen such that

(a) $P[T_n(\theta) \mid \theta] = 1 - \alpha$.

Then $T_n(\theta)$ satisfies also the following two conditions:

(b) The greatest lower bound of $P[T_n(\theta')\theta'']$ converges to $1 - \alpha$ with $n \to \infty$.

(c) For any sequence of regions $\{Z_n(\theta)\}$ which satisfies (a) and (b), the least upper bound of

$$P[Z_n(\theta') \mid \theta''] - P[T_n(\theta') \mid \theta'']$$

converges to 0 with $n \to \infty$.

[4] A. WALD, "Some examples of asymptotically most powerful tests," *Annals of Math. Stat.*, Vol 12 (1941), pp. 396–408.

On account of Propositions I and II we easily get the following theorems:

THEOREM I: *Denote by* $\xi_n(E_n)$ *the set of all values of* θ *for which* $y_n(\theta, E_n) \leq A_n(\theta)$ *and* $A_n(\theta)$ *is defined such that* $P[y_n(\theta, E_n) > A_n(\theta) \mid \theta] = 1 - \alpha$. *Then* $\xi_n(E_n)$ *satisfies the following two conditions:*

(a) $P[\xi_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(b) *For any sequence of setfunctions* $\{\xi'_n(E_n)\}$ *which satisfies the condition* (a), *the least upper bound of*

$$P[\xi_n(E_n) c \theta' \mid \theta''] - P[\xi'_n(E_n) c \theta' \mid \theta'']$$

in the set $\theta'' \geq \theta'$ *converges to* 0 *with* $n \to \infty$.

THEOREM II: *Denote by* $\zeta_n(E_n)$ *the set of all values of* θ *for which* $y_n(\theta, E_n) \geq B_n(\theta)$ *and* $B_n(\theta)$ *is defined such that* $P[y_n(\theta, E_n) < B_n(\theta) \mid \theta] = 1 - \alpha$. *Then* $\zeta_n(E_n)$ *satisfies the following two conditions:*

(a) $P[\zeta_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(b) *For any sequence of setfunctions* $\{\zeta'_n(E_n)\}$ *which satisfies the condition* (a), *the least upper bound of*

$$P[\zeta_n(E_n) c \theta' \mid \theta''] - P[\zeta'_n(E_n) c \theta' \mid \theta'']$$

in the set $\theta'' \leq \theta'$ *converges to* 0 *with* $n \to \infty$.

THEOREM III: *Denote by* $\rho_n(E_n)$ *the set of all values of* θ *for which* $| y_n(\theta, E_n) | \leq C_n(\theta)$ *and* $C_n(\theta)$ *is chosen such that* $P[| y_n(\theta, E_n) | > C_n(\theta) \mid \theta] = 1 - \alpha$. *Then* $\rho_n(E_n)$ *satisfies the following three conditions:*

(a) $P[\rho_n(E_n) c \theta \mid \theta] = \alpha$ *for all values of* θ.

(b) *The least upper bound of* $P[\rho_n(E_n) c \theta' \mid \theta'']$ *converges to* α *with* $n \to \infty$.

(c) *For any sequence of setfunctions* $\{\rho'_n(E_n)\}$ *which satisfies the conditions* (a) *and* (b), *the least upper bound of*

$$P[\rho_n(E_n) c \theta' \mid \theta''] - P[\rho'_n(E_n) c \theta' \mid \theta'']$$

converges to zero with $n \to \infty$.

Now we shall investigate the question whether the sets $\xi_n(E_n)$, $\zeta_n(E_n)$ and $\rho_n(E_n)$ are intervals. For this purpose we will prove some propositions.

PROPOSITION III: *Let* ϵ *and* D *be two positive numbers such that* $\epsilon < D$. *Denote by* $Q_n(\theta, \epsilon, D)$ *the region which consists of all points* E_n *for which*

$$y_n(\theta + \epsilon', E_n) \leq -n^{\frac{1}{4}}, \qquad \text{and} \qquad y_n(\theta - \epsilon', E_n) \geq n^{\frac{1}{4}}$$

for all values ϵ' *in the interval* $[\epsilon, D]$. *Then we have*

$$\lim_{n \to \infty} P[Q_n(\theta, \epsilon, D) \mid \theta] = 1 \tag{1}$$

uniformly in θ.

PROOF: Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_r$ be a sequence of points in the interval $[\epsilon, D]$ such that $\epsilon_1 - \epsilon = \epsilon_2 - \epsilon_1 = \cdots = \epsilon_r - \epsilon_{r-1} = D - \epsilon_r = k_0$ (say), where r is chosen sufficiently large such that Assumption II holds for $| \delta | \leq k_0$. Denote by $R_n(\theta, \epsilon_i)$ the region in which

$$y_n(\theta + \epsilon_i, E_n) \leq -n^{\frac{1}{4}}. \tag{2}$$

We will show that

$$(3) \qquad \lim_{n\to\infty} P[R_n(\theta, \epsilon_i) \mid \theta] = 1$$

uniformly in θ.

From Assumption I it follows that the greatest lower bound of

$$\left| E_\theta \frac{\partial}{\partial\theta} \log f(x, \theta + \epsilon') \right|$$

with regard to ϵ' in the interval $[\epsilon, D]$ is positive. Let this greatest lower bound be $A > 0$. Since on account of Assumption I $E_\theta \frac{\partial}{\partial\theta} \log f(x, \theta + \epsilon')$ is a continuous function of ϵ', it does not change sign in the interval $\epsilon \le \epsilon' \le D$. Since this is true for arbitrarily small ϵ and since $E_\theta \left[\frac{\partial}{\partial\theta} \log f(x,\theta)\right]^2 = -E_\theta \frac{\partial^2}{\partial\theta^2} \log f(x, \theta)$ has a positive lower bound (Assumption I), it follows easily on account of Assumption II that

$$E_\theta \frac{\partial}{\partial\theta} \log f(x, \theta + \epsilon') < 0.$$

Hence

$$(4) \qquad E_\theta \frac{\partial}{\partial\theta} \log f(x, \theta + \epsilon') \le -A < 0 \quad \text{for} \quad \epsilon \le \epsilon' \le D,$$

and therefore

$$(5) \qquad E_\theta y_n(\theta + \epsilon', E_n) \le -A\sqrt{n} \qquad \text{for } \epsilon \le \epsilon' \le D.$$

From Assumption II it follows that the variance of $y_n(\theta + \epsilon', E_n)$ is a bounded function of θ and ϵ'. Hence

$$(6) \qquad \lim_{n\to\infty} P[y_n(\theta + \epsilon_i, E_n) \le -\tfrac{1}{2}A\sqrt{n} \mid \theta] = 1$$

uniformly in θ. The equation (3) is a consequence of (6).

Denote by $S_n(\theta, \epsilon_i)$ the region in which

$$\left| \frac{1}{n} \sum_\alpha \varphi_i(x_\alpha, \theta + \epsilon_i, k_0) \right| < C \qquad (i = 1, 2)$$

where C is greater than the least upper bound of $|E_\theta\varphi_i(x, \theta', k_0)|$ with respect to θ and θ'. Then we have on account of Assumption II:

$$(7) \qquad \lim_{n\to\infty} P[S_n(\theta, \epsilon_i) \mid \theta] = 1 \qquad (i = 1, 2, \cdots, r)$$

uniformly in θ. In the region $S_n(\theta, \epsilon_i)$ we obviously have

$$(8) \qquad y_n(\theta + \epsilon_i', E_n) \le y_n(\theta + \epsilon_i, E_n) + 2k_0\sqrt{n}C$$

for all values ϵ_i' in the interval $[\epsilon_i - k_0, \epsilon_i + k_0]$. By choosing r sufficiently large we can always achieve that

$$2k_0 C \leq \frac{A}{4}.$$

Denote by $T_n(\theta, \epsilon_i)$ the region in which

$$y_n(\theta + \epsilon_i', E_n) \leq -\frac{A}{4}\sqrt{n} \quad \text{for} \quad \epsilon_i - k_0 \leq \epsilon_i' \leq \epsilon_i + k_0. \tag{9}$$

From (6), (7) and (8) we get

$$\lim_{n=\infty} P[T_n(\theta, \epsilon_i) \mid \theta] = 1 \tag{10}$$

uniformly in θ. Let $Q_n'(\theta, \epsilon, D)$ be the common part of the r regions $T_n(\theta, \epsilon_1), \cdots, T_n(\theta, \epsilon_r)$, i.e. $Q_n'(\theta, \epsilon, D)$ is the set of all points E_n for which

$$y_n(\theta + \epsilon', E_n) \leq -\frac{A}{4}\sqrt{n}$$

for all ϵ' in the interval $[\epsilon, D]$. Since r is a fixed positive integer not depending on n, we get from (10)

$$\lim_{n=\infty} P[Q_n'(\theta, \epsilon, D) \mid \theta] = 1 \tag{11}$$

uniformly in θ.

In the same way we can prove that

$$\lim_{n=\infty} P[Q_n''(\theta, \epsilon, D) \mid \theta] = 1 \tag{12}$$

uniformly in θ, where $Q_n''(\theta, \epsilon, D)$ denotes the region in which

$$y_n(\theta - \epsilon', E_n) \geq \frac{A}{4}\sqrt{n} \quad \text{for all} \quad \epsilon' \text{ in } [\epsilon, D].$$

Proposition III follows from (11) and (12).

PROPOSITION IV: *Denote by $V_n(\theta, \epsilon)$ the region in which*

$$\frac{\partial}{\partial\theta} y_n(\theta', E_n) < -n^{\frac{1}{2}}$$

for all values θ' in the interval $[\theta - \epsilon, \theta + \epsilon]$. There exists a positive ϵ such that

$$\lim_{n=\infty} P[V_n(\theta, \epsilon) \mid \theta] = 1$$

uniformly in θ.

PROOF: Since the least upper bound of $E_\theta\varphi_2(x, \theta, 0)$ is <0, we get from Assumption II that the least upper bound of $E_\theta\varphi_2(x, \theta, \epsilon)$ is <0 for sufficiently

small $\epsilon > 0$. Denote the least upper bound of $E_\theta \varphi_2(x, \theta, \epsilon)$ by $-B$ and let the region in which

$$\frac{1}{n} \sum_{\alpha} \varphi_2(x_\alpha, \theta, \epsilon) < -\tfrac{1}{2}B$$

be denoted by $W_n(\theta, \epsilon)$. From Assumption II it follows that

$$\lim_{n=\infty} P[W_n(\theta, \epsilon) \mid \theta] = 1$$

uniformly in θ. Since for almost all n $W_n(\theta, \epsilon)$ is a subset of $V_n(\theta, \epsilon)$, Proposition IV is proved.

Proposition V: *Let $A_n(\theta)$, $B_n(\theta)$, $C_n(\theta)$ be the functions as defined in Theorems I–III. There exists a finite value G such that*

$$| A_n(\theta) | < G, \qquad | B_n(\theta) | < G \quad \text{and} \quad | C_n(\theta) | < G$$

for all θ and all n.

Proposition V follows easily from the fact that the variance of $y_n(\theta, E_n)$ is a bounded function of n and θ.

Let D be an arbitrary positive number and denote by $W_n(\theta, D)$ the region consisting of all points E_n for which the following conditions are fulfilled:

(a) The equation $y_n(\theta', E_n) = A_n(\theta')$ has exactly one root in θ' which lies in the interval $[\theta - D, \theta + D]$.
(b) The equation $y_n(\theta', E_n) = B_n(\theta')$ has exactly one root in θ' which lies in the interval $[\theta - D, \theta + D]$.
(c) The equation $y_n(\theta', E_n) = C_n(\theta')$ has exactly one root in θ' which lies in the interval $[\theta - D, \theta + D]$.
(d) The equation $y_n(\theta', E_n) = -C_n(\theta')$ has exactly one root in θ' which lies in the interval $[\theta - D, \theta + D]$.
(e) The common part of $[\theta - D, \theta + D]$ and $\xi_n(E_n)$ is the interval $[\theta_n'(E_n), D]$ where $\theta_n'(E_n)$ denotes the root of the equation in (a).
(f) The common part of $\zeta_n(E_n)$ and $[\theta - D, \theta + D]$ is the interval $[-D, \theta_n''(E_n)]$ where $\theta_n''(E_n)$ denotes the root of the equation in (b).
(g) The common part of $\rho_n(E_n)$ and $[\theta - D, \theta + D]$ is the interval $[\underline{\theta}_n(E_n), \bar{\theta}_n(E_n)]$ where $\underline{\theta}_n(E_n)$ denotes the root of the equation in (c) and $\bar{\theta}_n(E_n)$ denotes the root of the equation in (d).

From Propositions III–V follows easily the following

Proposition VI. *For any positive value D*

$$\lim_{n=\infty} P[W_n(\theta, D) \mid \theta] = 1,$$

uniformly in θ, provided that the functions $A_n(\theta)$, $B_n(\theta)$ and $C_n(\theta)$ are continuous and of bounded variation in any finite interval.

We will show that Proposition VI remains valid for $D = +\infty$, if we make the following

ASSUMPTION V: *Denote by* $\psi(x, \theta, D)$ *the least upper bound of* $\frac{\partial}{\partial\theta}\log f(x, \theta')$ *with respect to* θ' *where* $\theta' \geq \theta + D$. *Denote furthermore by* $\psi^*(x, \theta, D)$ *the greatest lower bound of* $\frac{\partial}{\partial\theta}\log f(x,\theta')$ *with respect to* θ' *where* $\theta' \leq \theta - D$. *There exists a positive* D *such that the least upper bound of* $E_\theta\psi(x, \theta, D)$ *with respect to* θ *is negative, the greatest lower bound of* $E_\theta\psi^*(x, \theta, D)$ *with respect to* θ *is positive, and the variances of* $\psi(x, \theta, D)$ *and* $\psi^*(x, \theta, D)$ *are bounded functions of* θ. *(The variances are calculated under the assumption that* θ *is the true value of the parameter.)*

It follows easily from Assumption V that

$$\lim_{n=\infty} P\left[\frac{1}{\sqrt{n}}\sum_\alpha \psi(x_\alpha, \theta, D) < -n^{\frac{1}{4}} \mid \theta\right]$$
$$= \lim_{n=\infty} P\left[\frac{1}{\sqrt{n}}\sum_\alpha \psi^*(x_\alpha, \theta, D) > n^{\frac{1}{4}} \mid \theta\right] = 1$$

uniformly in θ.

Since

$$\frac{1}{\sqrt{n}}\sum_\alpha \psi(x_\alpha, \theta, D) \geq y_n(\theta', E_n) \quad \text{for} \quad \theta' \geq \theta + D$$

and

$$\frac{1}{\sqrt{n}}\sum_\alpha \psi^*(x_\alpha, \theta, D) \leq y_n(\theta', E_n) \quad \text{for} \quad \theta' \leq \theta - D,$$

Proposition VI remains valid if we substitute $+\infty$ for D.

Hence we obtain the following

COROLLARY: *If the assumptions* I–V *are fulfilled and if* $A_n(\theta)$, $B_n(\theta)$ *and* $C_n(\theta)$ *are continuous and of bounded variation in any finite interval, then*

(a) *The root* $\theta_n'(E_n)$ *of the equation* $y_n(\theta, E_n) = A_n(\theta)$ *in* θ *is an asymptotically best lower estimate of* θ.

(b) *The root* $\theta_n''(E_n)$ *of the equation* $y_n(\theta, E_n) = B_n(\theta)$ *in* θ *is an asymptotically best upper estimate of* θ.

(c) *The interval* $[\underline{\theta}_n(E_n), \bar{\theta}_n(E_n)$ *is an asymptotically shortest unbiased confidence interval of* θ, *where* $\underline{\theta}_n(E_n)$ *denotes the root of the equation* $y_n(\theta, E_n) = +C_n(\theta)$, *and* $\bar{\theta}_n(E_n)$ *denotes the root of the equation* $y_n(\theta, E_n) = -C_n(\theta)$.

5. Some Remarks. 1. I should like to make a few remarks about the relationship of these results to those obtained by S. S. Wilks.[5] The definition of a shortest confidence interval underlying Wilks' investigations is somewhat different from that of Neyman's which has been used in this paper. According to Wilks, a confidence interval $\delta(E_n)$ is called shortest in the average if the expected

[5] S. S WILKS, "Shortest average confidence intervals from large samples," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 166–175.

value of the length of $\delta(E_n)$ is a minimum. The main result obtained by Wilks can be formulated as follows: The confidence interval $[\underline{\theta}_n(E_n), \bar{\theta}_n(E_n)]$ given in our Corollary is asymptotically shortest in the average compared with all confidence intervals computed on the basis of functions belonging to a certain class C. In the present paper no restriction to a certain class of functions has been made.

2. If the parameter space Ω is not the whole real axis, but an open subset of it, and if the assumptions I–V are fulfilled when θ can take only values in Ω, the previously proved Corollary remains valid. If Ω is a bounded set, Assumption V is a consequence of Assumptions I–IV.

GROUPING METHODS

BY P. S. DWYER

University of Michigan

1. Introduction. The conventional formulas for moment adjustments known as Sheppard's corrections are not too satisfactory for practical use. As Carver has pointed out [1] Sheppard's corrections are merely systematic adjustments which eliminate the bias introduced by grouping. The values of the moments after Sheppard's corrections have been applied may be looked upon as unbiased grouping estimates of the true moments while the uncorrected values constitute biased estimates

In practice one obtains his moments from a single grouping. The application of Sheppard's adjustments in such a case does not necessarily result in the unbiased estimate being closer to the true moment than is the biased estimate and, in an appreciable percentage of cases, the unbiased estimate is further from the true moment than is the biased estimate. One does not know when he applies Sheppard's adjustments to the results of a single grouping whether or not he is making a correction in the right direction.

This situation is not too satisfactory and yet practical necessity demands some method of grouping The improvement of modern calculating machines tends to push grouping techniques further into the background since, in many cases, the machines permit the determination of the actual values of the moments without grouping in a reasonable amount of time. But even here it is possible to use grouping methods and to get a good estimate of the true value in a fraction of the time. It is the purpose of this paper to present some new grouping methods which are useful in obtaining much better unbiased estimates from a single grouping than can be obtained with the use of Sheppard's corrections. These methods demand additional work but this additional work is justified by the additional precision resulting when such precision is desired.

The spirit of the new approach, which in one sense is a generalization of the earlier approach, can be expressed very simply though the details of the development and the calculational methods demand amplification. If we let x = the true value and x' the grouped value (the value of the class mark of the group in which x is), and the error, ϵ = the difference between the true value and the grouped value, then

$$(1) \qquad \epsilon = x - x', \qquad x = x' + \epsilon, \quad \text{and} \quad x' = x - \epsilon.$$

In the classical theory we use $\Sigma x'^s$ as the biased grouping estimate of Σx^s. In the new methods we use $\Sigma x x'^{s-1}$ as the biased estimate or, if we desire more precision, $\Sigma x^2 x'^{s-2}$ as the biased estimate. It is then possible to correct $\Sigma x x'^{s-1}$ for grouping bias and to correct $\Sigma x^2 x'^{s-2}$ for grouping bias just as we now correct $\Sigma x'^s$ for grouping bias. It is also possible to use the values of $\Sigma x'^s$ and $\Sigma x x'^{s-1}$

in obtaining a better unbiased estimate or to use the values of $\Sigma x'^s$, $\Sigma x x'^{s-1}$, and $\Sigma x^2 x'^{s-2}$ in obtaining a still better unbiased estimate of Σx^s.

2. Illustration. The relative merits of the conventional method and the proposed methods can be shown effectively by means of an illustration. For this purpose I have selected the problem used previously [1, 154] in showing the variations in grouped results. The power sums rather than the moments are used and the origin is taken at a point near the mean so that the relative variations are as large as possible. If the values of the power sums were "padded" by measurement about zero, the relative variations would not appear as large. However, a problem which shows considerable variation, and in the problem under consideration the nine $\Sigma x'^3$ unbiased grouping estimates of Σx^3 resulting from the nine groupings do not even have the same sign, is an appropriate one with which to demonstrate the improvements introduced by the new methods.

The problem consists of 244 discrete variates which range in value from 64 to 155 Carver took a class interval of nine and formed the nine frequency distributions which result when class intervals of nine are chosen in all possible ways. He computed the values of $\Sigma x'$, $\Sigma x'^2$, $\Sigma x'^3$, $\Sigma x'^4$, for each of the nine distributions, corrected each for bias with the use of the Sheppard adjustments, and showed that the averages of the nine corrected estimates are respectively the values of Σx, Σx^2, Σx^3, Σx^4.

In Table I are presented the values of the biased and unbiased grouping estimates of Σx, Σx^2, Σx^3, Σx^4, which result from the use of (1) $\Sigma x'^s$; (2) $\Sigma x x'^{s-1}$; (3) $\Sigma x^2 x'^{s-2}$; (4) $\Sigma x'^s$, $\Sigma x x'^{s-1}$; and (5) $\Sigma x'^s$, $\Sigma x x'^{s-1}$, $\Sigma x^2 x'^{s-2}$. The results are presented here for comparison only; the details of the computation are explained later. Rows of biased estimates are indicated by B while the rows of unbiased estimates are indicated by U. Parentheses are used to indicate entries which, while appearing in rows of biased estimates, are actually unbiased. The exact values of Σx^s, when they appear, are indicated by underscoring. The Roman numerals indicate the different frequency distributions while the grouping methods are indicated by the values of (1), (2), (3), (4), (5) above The true values are $\Sigma x = -129$, $\Sigma x^2 = 77{,}591$, $\Sigma x^3 = -52{,}005$, $\Sigma x^4 = 69{,}239{,}951$ where the values of x used are the values of the original variates decreased by 105.

The information contained in Table I deserves more than cursory examination. Study shows that the estimates resulting from method 2 are much closer to the true values than are the estimates resulting from method 1, etc. Table II is presented below in order to facilitate the comparison of the relative amounts of grouping error involved in the different methods. The standard deviation of the grouping error of the conventional method, method 1, is used as a norm and the standard deviations of the grouping errors for the new methods are compared with this norm.

The decline in the size of the error revealed in Table II indicates a decided decrease in grouping errors. Grouping method 2 enables one to compute the mean exactly and this is always possible when method 2 is applied to discrete

TABLE I

Biased and Unbiased Grouping Estimates by Different Methods

Grouping	Grouping Method	Σx -129	Σx^2 77,591	Σx^3 $-52,005$	Σx^4 69,219,951
I	1-B	(−181)	77,149	−134,101	69,003,265
I	2-B	(−129)	(76,593)	−105,105	68,267,577
I	3-B	—	(77,591)	(−77,825)	68,033,657
I	1-U	−181	75,522¼	−130,571	66,023,177
I	2-U	−129	76,593	−104,245	66,735,717
I	3-U	—	77,591	−77,825	67,516,383¾
I	4-U	−129	77,663¾	−49,513	69,001,817
I	5-U	−129	77,591	−52,351	69,193,537
II	1-B	(−218)	78,466	−54,602	74,519,062
II	2-B	(−129)	(77,181)	−52,977	72,296,367
II	3-B	—	(77,591)	(−52,465)	70,685,801
II	1-U	−218	76,839¼	−50,242	71,427,194
II	2-U	−129	77,181	−52,117	70,752,747
II	3-U	—	77,591	−52,465	70,168,527¼
II	4-U	−129	77,522¾	−52,307	68,770,106
II	5-U	−129	77,591	−53,066	69,246,172
III	1-B	(−111)	77,769	2,889	71,165,409
III	2-B	(−129)	(76,797)	−17,037	70,053,093
III	3-B	—	(77,591)	(−35,097)	69,211,545
III	1-U	−111	76,142¼	5,109	68,100,521
III	2-U	−129	76,797	−16,177	68,517,153
III	3-U	—	77,591	−35,097	68,697,271¾
III	4-U	−129	77,451¾	−59,469	68,945,609
III	5-U	−129	77,591	−51,291	69,201,169
IV	1-B	(−139)	79,747	−23,311	74,171,443
IV	2-B	(−129)	(77,790)	−34,464	72,095,064
IV	3-B	—	(77,591)	(−44,108)	70,502,774
IV	1-U	−139	78,120¼	−20,531	71,027,435
IV	2-U	−129	77,790	−33,604	70,539,864
IV	3-U	—	77,591	−44,108	70,075,500¾
IV	4-U	−129	77,459¾	−59,350	69,037,511
IV	5-U	−129	77,591	−52,243	69,248,077
V	1-B	(−104)	81,934	19,666	76,143,874
V	2-B	(−129)	(78,891)	−4,521	73,590,207
V	3-B	—	(77,591)	(−28,387)	71,010,053
V	1-U	−104	80,307¼	21,746	72,912,380
V	2-U	−129	78,891	−3,661	72,012,387
V	3-U	—	77,591	−28,387	71,092,779¾
V	4-U	−129	77,474¾	−55,475	69,142,430
V	5-U	−129	77,591	−51,932	69,312,700

TABLE I (*Cont'd.*)

Grouping	Grouping Method	Σx −129	Σx^2 77,591	Σx^3 −52,005	Σx^4 69,239,951
VI	1-B	(−87)	80,145	16,551	72,467,541
VI	2-B	(−129)	(78,030)	−4,914	70,940,124
VI	3-B	—	(77,591)	(−27,714)	69,902,910
VI	1-U	−87	78,518$\frac{1}{3}$	18,291	69,307,613
VI	2-U	−129	78,030	−4,054	69,379,524
VI	3-U	—	77,591	−27,714	69,385,636$\frac{2}{3}$
VI	4-U	−129	77,541$\frac{2}{3}$	−50,424	69,536,657
VI	5-U	−129	77,591	−51,849	69,241,507
VII	1-B	(−52)	80,302	−36,118	71,851,930
VII	2-B	(−129)	(78,168)	−39,486	70,515,354
VII	3-B	—	(77,591)	(−44,492)	69,647,462
VII	1-U	−52	78,675$\frac{1}{3}$	−35,078	68,685,722
VII	2-U	−129	78,168	−38,626	68,951,994
VII	3-U	—	77,591	−44,492	69,130,188$\frac{2}{3}$
VII	4-U	−129	77,660$\frac{2}{3}$	−48,802	69,689,930
VII	5-U	−129	77,591	−51,136	69,260,146
VIII	1-B	(−89)	78,553	−101,357	68,426,497
VIII	2-B	(−129)	(77,352)	−82,416	67,959,816
VIII	3-B	—	(77,591)	(−65,788)	67,944,476
VIII	1-U	−89	76,926$\frac{1}{3}$	−99,577	65,330,249
VIII	2-U	−129	77,352	−81,556	66,412,776
VIII	3-U	—	77,591	−65,788	67,427,202$\frac{2}{3}$
VIII	4-U	−129	77,777$\frac{2}{3}$	−47,114	69,711,437
VIII	5-U	−129	77,591	−51,473	69,210,235
IX	1-B	(−180)	78,894	−180,792	73,155,150
IX	2-B	(−129)	(77,517)	−134,865	71,407,737
IX	3-B	—	(77,591)	(−92,169)	70,183,341
IX	1-U	−180	77,267$\frac{1}{3}$	−177,192	70,045,262
IX	2-U	−129	77,517	−134,005	69,857,397
IX	3-U	—	77,591	−92,169	69,666,666$\frac{2}{3}$
IX	4-U	−129	77,766$\frac{2}{3}$	−45,591	69,323,762
IX	5-U	−129	77,591	−52,704	69,246,016

data. There is also a corresponding decrease in the errors of the higher powers to roughly one-half, two-thirds, three-fourths. Greater precision in the case of the higher power sums can be obtained with the use of the other methods, though these methods demand more calculation.

There is one more question which should be discussed before the general

theory is presented, and that deals with the method of computation of the quantities $\Sigma xx'^{s-1}$, $\Sigma x^2x'^{s-2}$, in methods 2 and 3. Computational techniques are discussed in a later section of the paper, but enough should be given now to make the meaning of $\Sigma xx'^{s-1}$ and $\Sigma x^2x'^{s-2}$ clear. In getting $\Sigma x'^s$, we recall, we need only the values of the class mark, x' and the frequency associated with each, $f_{x'}$. To get the values of $\Sigma xx'^{s-1}$ we need in addition to x' the sum of the x values which are grouped together in the class having the same class mark, x'. We denote this value by $x_{x'}$ and we use this instead of the $f_{x'}$ of the usual method. In the case of method 3 we record $x^2_{x'}$ where $x^2_{x'}$ is the sum of the squares of all x values having the same grouped value x'.

Let us examine the first grouping in Table I. The original 244 variates were recorded by Carver [1, 154] and he gave the values of $f_{x'}$ for each grouping. It is necessary for us to return to these original variates, but instead of counting the variates in a given group, we add them and we add their squares.

In obtaining the values for the first grouping in Table I the variates were transformed with the use of $x = v - 105$. The variates then ranged from

TABLE II

Standard Deviations of the Grouping Errors of the Different Methods Expressed as Percentages of the Standard Deviations of the Usual Method

Method	Σx	Σx^2	Σx^3	Σx^4
1	100	100	100	100
2	0	48.9	65.5	74.3
3	—	0	32.1	49.1
4	0	8.8	7.3	13.8
5	0	0	0	1.5

-41 to 50 and the frequency distribution was made with mid values $x' = -37$, -28, -19, -10, etc. The values of $f_{x'}$, $x_{x'}$, and $x^2_{x'}$ were then computed and recorded in the columns, 2, 3, 4, of Table III. The next three columns are computational columns useful in obtaining the biased estimates recorded at the bottom of Table III and also in Table I with the use of $\sum_{x'} x'^s f_{x'}$,

$$\sum xx'^{s-1} = \sum_{x'} x_{x'}x'^{s-1}, \qquad \sum x^2x'^{s-2} = \sum_{x'} x^2_{x'}x'^{s-2}.$$

3. General formulas for corrections for grouping bias. We are next led to the question of correcting these estimates of Σx^s for the bias introduced by grouping. Before indicating the numerical work, we derive general formulas for correction for grouping bias.

We assume that the variates are recorded in units of h which means that, in the case of discrete series, the smallest possible difference between any two unequal variates is equal to h. In case the distribution is continuous, the recorded values constitute a discrete series recorded in units of h. Thus heights may be recorded to the nearest inch, in which case h is one inch, or to the nearest

one hundredth of an inch, in which case h is 1/100 inch, etc. We assume further that all possible groupings of k different values are made. Thus if the smallest variate is a, then the values of $x = a, a + h, a + 2h, \cdots, a + \overline{k + i}h, \cdots, a + \overline{k-1}\, h$ are thrown in a group with class mark $a + \frac{1}{2}(k - 1)h$. The k possible sets of groupings of k are made in this way.

We examine the error involved when a specific variate x is replaced by the class mark x' in each of these groupings. The values of the lower open limit, L, the upper open limit, U, the class mark, x', and the error $\epsilon = x - x'$ are indicated in Table IV. The k different groupings indicated by the different rows show x at the lower limit, x one step above the lower limit, x two steps above

TABLE III

Values of x', $f_{x'}$, $x_{x'}$ and $x^2_{x'}$ for the First Grouping with Computation of Biased Estimates of $\Sigma x'^t$

x' \ F_x	$f_{x'}$	$x_{x'}$	$x^2_{x'}$	x'^2	x'^3	x'^4
53	1	50	2500	2809	148,877	7,890,481
44	1	48	2304	1936	85,184	3,748,096
35	8	287	10351	1225	42,875	1,500,625
26	16	402	10190	676	17,576	456,976
17	27	475	8515	289	4,913	83,521
8	45	378	3426	64	512	4,096
−1	53	−73	369	1	−1	1
−10	41	−386	3924	100	−1000	10,000
−19	27	−507	9701	361	−6859	130,321
−28	12	−338	9594	784	−21,952	614,656
−37	13	−465	16717	1369	−50,653	1,874,161
ΣF_x	244	−129	77591			
$\Sigma x' F_x$	−181	76,593	−77825			
$\Sigma x'^2 F_x$	77,149	−105,105	68,033,657			
$\Sigma x'^3 F_x$	−134,191	68,267,577				
$\Sigma x'^4 F_x$	69,063,265					

the lower limit, etc. It is at once apparent that the errors in replacing x' for x in the k different ways constitute the deviations from the mean of the rectangular distribution $h, 2h, 3h, \cdots, kh$. We indicate the moments about the mean of this rectangular distribution by R_1, R_2, R_3, R_4 and we use the notation $E(\epsilon^t)$ as the sum of the tth powers of the k different ϵ's divided by k. It follows that $E(\epsilon^t) = R_t$. Now the values of R_t are 0 when t is odd and are well known when t is even [2,325]. The ones in which we are especially interested are

$$R_2 = \frac{k^2 - 1}{12} h^2 \quad \text{and} \quad R_4 = \frac{(k^2 - 1)(3k^2 - 7)}{240} h^4. \tag{2}$$

If an adjustment of scale is made so that the differences between successive class marks are unity, as is customary, the value of h is $1/k$. The values of R_2 and R_4 are then

$$(3) \qquad R_2 = \frac{1 - 1/k^2}{12}, \qquad R_4 = \frac{(1 - 1/k^2)(3 - 7/k^2)}{240}.$$

As the number of groupings increases the value of $1/k^2 \to 0$ and the appropriate values of the moments of the continuous rectangular distribution result. Thus

$$R_2 = \frac{1}{12}, \qquad R_4 = \frac{1}{80}, \quad \text{and} \quad 6R_2^2 - R_4 = \frac{7}{240}.$$

TABLE IV

Open Limits, Class Marks and Errors for the Different Groupings

Grouping	x	$L =$	$U =$	$x' = \frac{1}{2}(L + U)$	$\epsilon = x - x'$
1	$x = L$	x	$x + (k-1)h$	$x + \frac{1}{2}(k-1)h$	$-\frac{1}{2}(k-1)h$
2	$x = L + h$	$x - h$	$x + (k-2)h$	$x + \frac{1}{2}(k-3)h$	$-\frac{1}{2}(k-3)h$
3	$x = L + 2h$	$x - 2h$	$x + (k-3)h$	$x + \frac{1}{2}(k-5)h$	$-\frac{1}{2}(k-5)h$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	$x = L + \overline{i-1}h$	$x - \overline{i-1}h$	$x + (k-i)h$	$x + \frac{1}{2}[k-(2i-1)]h$	$-\frac{1}{2}[k-(2i-1)]h$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$k-1$	$x = L + (k-2)h$	$x - (k-2)h$	$x + h$	$x - \frac{1}{2}(k-3)h$	$\frac{1}{2}[k-3]h$
k	$x = U = L + (k-1)h$	$x - (k-1)h$	x	$x - \frac{1}{2}(k-1)h$	$\frac{1}{2}[k-1]h$

If now we let F_x be any real function of x defined for the values $x = a, a + k, a + 2k, \cdots$, we have at once the useful lemma

$$(4) \qquad E[\Sigma x^a \epsilon^t F_x] = \Sigma x^a F_x E[\epsilon^t] = R_t \Sigma x^a F_x.$$

This results from the fact that the values of x, and of all functions of x, are unchanged by the groupings even though the values of x' and ϵ vary.

The Σ in (4) indicates a summation with respect to the variates while the summation with respect to the different errors is taken care of in the E notation. The limits of the Σ in (4) are purposely left indefinite so that either a serial or a frequency notation can be used. Thus if a serial notation is used, the limits are from 1 to N, the values of x are the variates x_i and F_x becomes F_{x_i}. In this case F_{x_i} may be set equal to unity to give the corrections of method 1, may be set equal to x_i to give the corrections of method 2, or may be set equal

to x_i^2 to give the corrections of method 3. In case the notation of the frequency distribution is preferred, the limits of the summation are the smallest variate and the largest variate, the values of x are the values of the different variates which occur. In this case we may have $F_x = f_{x'}$, the frequency function, $F_x = xf_{x'} = x_{x'}$, or $F_x = x^2f_{x'} = x^2_{x'}$.

The continued application of (4) to the terms in the expansion of $E[\Sigma x'^s F_x]$ results in

$$(5)\qquad \begin{aligned} E\left[\sum x'^s F_x\right] &= E\left[\sum (x - \epsilon)^s F_x\right] = E\left[\sum_{t=0}^{s} (-1)^t \binom{s}{t} x^{s-t}\epsilon^t F_x\right] \\ &= \sum_{t=0}^{s} (-1)^t \binom{s}{t} \sum x^{s-t} F_x E(\epsilon^t) = \sum_{t=0}^{s} (-1)^t \binom{s}{t} R_t \sum x^{s-t} F_x . \end{aligned}$$

The fact that $R_t = 0$ when t is odd may be used in writing out the expansion. It is possible to work out a more general theory where the class mark is some other value (say the smallest variate) rather than the mid-value. In such a case formula (5) would apply, but the values of R_t would be the values of the moments of a rectangular distribution rather than the central moments. The above formula is sufficiently general for the purposes of this paper.

Specific values of (5) when $s = 0, 1, 2, 3, 4$ are

$$(6)\qquad \begin{aligned} E[\Sigma F_x] &= \Sigma F_x \\ E[\Sigma x' F_x] &= \Sigma x F_x \\ E[\Sigma x'^2 F_x] &= \Sigma x^2 F_x + R_2 \Sigma F_x \\ E[\Sigma x'^3 F_x] &= \Sigma x^3 F_x + 3R_2 \Sigma x F_x \\ E[\Sigma x'^4 F_x] &= \Sigma x^4 F_x + 6R_2 \Sigma x^2 F_x + R_4 \Sigma F_x . \end{aligned}$$

These equations can be solved for ΣF_x, $\Sigma x F_x$, etc., in terms of the expected values. If we use the inverse operator and write $E^{-1}[B] = A$ instead of $E[A] = B$ we have

$$(7)\qquad \begin{aligned} E^{-1}[\Sigma F_x] &= \Sigma F_x \\ E^{-1}[\Sigma x F_x] &= \Sigma x' F_x \\ E^{-1}[\Sigma x^2 F_x] &= \Sigma x'^2 F_x - R_2 \Sigma F_x \\ E^{-1}[\Sigma x^3 F_x] &= \Sigma x'^3 F_x - 3R_2 \Sigma x' F_x \\ E^{-1}[\Sigma x^4 F_x] &= \Sigma x'^4 F_x - 6R_2 \Sigma x'^2 F_x + (6R_2^2 - R_4)\Sigma F_x \text{ and in general,} \end{aligned}$$

$$E^{-1}\left[\sum x^s F_x\right] = \sum x'^s F_x - \sum_{t=2}^{s} (-1)^t \binom{s}{t} R_t E^{-1}\left[\sum x^{s-t} F_x\right].$$

These values $E^{-1}[\Sigma x^s F_x]$ are unbiased estimates of $\Sigma x^s F_x$ since

$$E[E^{-1}[\Sigma x^s F_x]] = \Sigma x^s F_x$$

The corrections for method 1, the customary corrections, are obtained if a serial notation is used with $F_x = 1$. The corrections for method 2 are obtained if a serial notation is used with $F_x = x$. The corrections for method 3 are obtair d with $F_x = x^2$. Thus we have

$$
\begin{aligned}
&E^{-1}[N] = N \\
&E^{-1}[\Sigma x] = \Sigma x' \\
(8) \qquad &E^{-1}[\Sigma x^2] = \Sigma x'^2 - R_2 N \\
&E^{-1}[\Sigma x^3] = \Sigma x'^3 - 3R_2 \Sigma x' \\
&E^{-1}[\Sigma x^4] = \Sigma x'^4 - 6R_2 \Sigma x'^2 + (6R_2^2 - R_4)N;
\end{aligned}
$$

$$
\begin{aligned}
&E^{-1}[\Sigma x\,] = \Sigma x \\
&E^{-1}[\Sigma x^2] = \Sigma x x' \\
(9) \qquad &E^{-1}[\Sigma x^3] = \Sigma x x'^2 - R_2 \Sigma x \\
&E^{-1}[\Sigma x^4] = \Sigma x x'^3 - 3R_2 \Sigma x x' \quad \text{etc.,}
\end{aligned}
$$

and

$$
\begin{aligned}
&E^{-1}[\Sigma x^2] = \Sigma x^2 \\
(10) \qquad &E^{-1}[\Sigma x^3] = \Sigma x^2 x' \\
&E^{-1}[\Sigma x^4] = \Sigma x^2 x'^2 - R_2 \Sigma x^2.
\end{aligned}
$$

These formulas are the ones used in obtaining the unbiased estimates in methods 1, 2, 3 from the biased estimates in Table I. In this case $R_2 = \dfrac{9^2 - 1}{12} = \dfrac{20}{3}$, $6R_2^2 - R_4 = \dfrac{(9^2 - 1)(7 \cdot 9^2 - 3)}{240} = 188$, $N = 244$ and the values follow by direct substitution in (8), (9), (10) above.

4. Compound grouping formulas. So far nothing has been said about the calculation of the results by methods 4 and 5. These methods might be called compound grouping methods, since they utilize the biased results of more than one grouping method. The values of $\Sigma x'^s$ and $\Sigma x x'^{s-1}$ are needed for method 4 and the values of $\Sigma x'^s$, $\Sigma x x'^{s-1}$, $\Sigma x^2 x'^{s-2}$ for method 5. The formulas for method 4 are first presented. The argument is given in some detail for the value of $E^{-1}[\Sigma x^2]$. Now

$$
\begin{aligned}
\Sigma x^2 = \Sigma(x' + \epsilon)^2 &= \Sigma x'^2 + 2\Sigma x'\epsilon + \Sigma \epsilon^2 \\
&= \Sigma x'^2 + 2\Sigma x'(x - x') + \Sigma \epsilon^2
\end{aligned}
$$

so that

$$\Sigma x^2 = -\Sigma x'^2 + 2\Sigma x x' + \Sigma \epsilon^2.$$

If the values of ϵ are known, we would have the exact value of Σx^2 since we know $\Sigma x'^2$ and $\Sigma xx'$. However, we do not know these values of ϵ from a single grouping, so we derive a formula giving unbiased estimates of Σx^2. We have at once

$$E[\Sigma x^2] = \Sigma x^2 = E[-\Sigma x'^2 + 2\Sigma xx'] + NR_2$$

and since $NR_2 = E[NR_2]$ we have

$$\text{(11)} \qquad \begin{aligned} \Sigma x^2 &= E[-\Sigma x'^2 + 2\Sigma xx' + NR_2] \quad \text{and} \\ E^{-1}[\Sigma x^2] &= -\Sigma x'^2 + 2\Sigma xx' + NR_2 . \end{aligned}$$

There is a relatively small error in this estimate since the only error involved is the difference between NR_2 and the actual sum of the squares of the ϵ's. This formula is the basis of the values of $E^{-1}[\Sigma x^2]$ recorded in Table I under method 4. For example, in grouping I, the estimate is $-77149 + 2(76593) + 244(\frac{20}{3}) = 77663\frac{2}{3}$ and this differs by only $72\frac{2}{3}$ from the exact value.

In a corresponding manner, we may prove

$$\text{(12)} \qquad \begin{aligned} E^{-1}[\Sigma x^3] &= -2\Sigma x'^3 + 3\Sigma xx'^2 + 3R_2\Sigma x \\ E^{-1}[\Sigma x^4] &= -3\Sigma x'^4 + 4\Sigma xx'^3 + 6R_2E^{-1}[\Sigma x^2] + 3NR_4 . \end{aligned}$$

Different values of $E^{-1}[\Sigma x^2]$ can be used. In the calculations of Table I the values $E^{-1}[\Sigma x^2] = \Sigma xx'$ from (9) were used, but the values $E^{-1}[\Sigma x^2] = -\Sigma x'^2 + 2\Sigma xx' + NR_2$ could be used to give somewhat better results.

It can be shown also that

$$\text{(13)} \qquad \begin{aligned} E^{-1}[\Sigma x^5] &= -4\Sigma x'^5 + 5\Sigma xx'^4 + 10R_2E^{-1}[\Sigma x^3] + 15R_4\Sigma x; \\ E^{-1}[\Sigma x^6] &= -5\Sigma x'^6 + 6\Sigma xx'^5 + 15R_2E^{-1}[\Sigma x^4] + 45R_4E^{-1}[\Sigma x^2] + 5R_6\mathrm{N}, \end{aligned}$$

and, after some argument that

$$\text{(14)} \qquad \begin{aligned} E^{-1}[\textstyle\sum x^s] = -(s-1)\textstyle\sum x'^s + s\textstyle\sum x'^{s-1}x \\ + \sum_{t=1}^{[\frac{1}{2}s]} \binom{s}{2t}(2t-1)R_{2t}E^{-1}\left[\textstyle\sum x^{s-2t}\right]' \end{aligned}$$

where $[\frac{1}{2}s]$ indicates the integer $\frac{1}{2}s$ or $\frac{1}{2}(s-1)$.

It is possible to obtain better unbiased estimates if we use in addition the values of $\Sigma x^2x'^{s-2}$. In this case the values of Σx and Σx^2 are known exactly, and after expansion of $\Sigma(x' + \epsilon)^s$, replacement of ϵ by $x - x'$ and of ϵ^2 by $(x - x')^2$, and further reduction, we get

$$\text{(15)} \qquad \begin{aligned} E^{-1}[\Sigma x^3] &= \Sigma x'^3 - 3\Sigma xx'^2 + 3\Sigma x^2x', \\ E^{-1}[\Sigma x^4] &= 3\Sigma x'^4 - 8\Sigma xx'^3 + 6\Sigma x^2x'^2 - 3NR_4 , \\ E^{-1}[\Sigma x^5] &= 6\Sigma x'^5 - 15\Sigma xx'^4 + 10\Sigma x^2x'^3 - 15R_4\Sigma x , \\ E^{-1}[\Sigma x^6] &= 10\Sigma x'^6 - 24\Sigma xx'^5 + 15\Sigma x^2x'^4 - 45R_4\Sigma x^2 - 10NR_6 , \end{aligned}$$

and in general,

$$E^{-1}\left[\sum x^s\right] = \tfrac{1}{2}(s-1)(s-2)\sum x'^s - s(s-2)\sum x x'^{s-1} \tag{16}$$
$$+ \tfrac{1}{2}s(s-1)\sum x^2 x'^{s-2} - \sum_{t=2}^{[s/2]} \binom{s}{2t}\binom{2t-1}{2} R_{2t} E^{-1}\left[\sum x^{s-2t}\right].$$

Compound formulas involving additional quantities such as $\Sigma x^3 x'^{s-3}$, $\Sigma x^4 x'^{s-4}$, etc., can be worked out by the methods outlined above.

5. Computational methods. It has been shown in sections 3 and 4 how the unbiased estimates can be obtained from the biased estimates. It is the purpose of the present section to show how these biased estimates can be computed efficiently. One method of calculation was shown in Table III. The values of $f_{x'}$, $x_{x'}$, $x^2_{x'}$ were computed and recorded, and the resulting power sums obtained. This is the most direct means of computation and if the number of groups is small and if a modern computing machine equipped with automatic positive and negative multiplication is available, it may be the preferred method. It should be noted that the values of $x^2_{x'}$ in Table III are obtained most easily with the use of a machine which permits the calculation of the square with a single key punching operation.

It is customary to use the devices of subtraction of a constant (either a central class mark or the smallest class mark) and division by a constant (size of the class interval) to simplify the computational work. Thus in Table III we could use the transformation $d' = \dfrac{x' - (-37)}{9}$ and compute the values of $\Sigma d'^s F_{x'}$. If F_x is the frequency function, we have the usual formulas, but if F_x is $x_{x'}$ or $x^2_{x'}$, then the results are terms of the type $\Sigma x d'^{s-1}$ or $\Sigma x^2 d'^{s-2}$. It is possible to reduce these to equivalent variables by the use of $d = \dfrac{x - (-37)}{9}$ so that the values of $\Sigma d'^s$, $\Sigma d d'^{s-1}$, $\Sigma d^2 d'^{s-2}$ result. We then correct for bias with the use of the formulas of sections 3 and 4 where the power sums of the rectangular distribution are computed with $h = 1/k$.

Another method which in many cases is preferable to that just described is the method of cumulative totals. The values of $f_{x'}$, $x_{x'}$, and $x^2_{x'}$, are cumulated successively for the different values of x' and the values of the biased grouping estimates are obtained immediately from the entries in the last few rows. The cumulations of Table III are shown in Table V. The entries in the column of the highest cumulations of $f_{x'}$, $x_{x'}$, $x^2_{x'}$, with the exception of those at the bottom of the column, need not be recorded.

It is possible to provide multipliers for these entries by an adaptation of a method given in an earlier paper [3]. A table of multipliers has a top marginal row composed of a, $a + k$, $a + 2k$, etc., and a left marginal column composed of $k - a$, $2k - a$, etc. The first row in the table is composed of 1, $k - a$, $(k - a)^2$, $(k - a)^3$, etc., and the first column of 1, a, a^2, a^3, etc. Each entry in the table is found by adding the product of the entry above it and the columnar heading

TABLE V

Computation of Biased Estimates Using Cumulative Totals

x'	$C(f_{x'})$	$C^2(f_{x'})$	$C^3(f_{x'})$	$C^4(f_{x'})$	$C^5(f_{x'})$	$C(x_{x'})$	$C^2(x_{x'})$	$C^3(x_{x'})$	$C^4(x_{x'})$	$C(x^2_{x'})$	$C^2(x^2_{x'})$	$C^3(x^2_{x'})$
53	1	1	1	1	—	50	50	50	—	2,500	2,500	—
44	2	3	4	5	—	98	148	198	—	4,804	7,304	—
35	10	13	17	22	—	385	533	731	—	15,155	22,459	—
26	26	39	56	78	—	787	1,320	2,051	—	25,345	47,804	—
17	53	92	148	226	—	1,262	2,582	4,633	—	33.860	81,664	—
8	98	190	338	564	—	1,640	4,222	8,855	—	37,286	118,950	—
−1	151	341	679	1,243	2,139	1,567	5,789	14,644	—	37,655	156,605	—
−10	192	533	1,212	2,455	4,594	1,181	6,970	21,614	52,776	41,579	198,184	—
−19	219	752	1,964	4,419	9,013	674	7,644	29,258	82,034	51,280	249,464	884,934
−28	231	983	2,947	7,366	16,379	336	7,980	37,238	119,272	60,874	310,338	1,195,272
−37	244	1227	4,174	11,540	27,919	−129	7,851	45,089	164,361	77,591	387,929	1,583,201
Σ	244	−181	77,149	−134,191	69,063,265	−129	76,593	−105,105	68,267,577	77,591	−77,825	68,033,657

to the product of the entry at the left and the row heading. The multipliers for $a = -37$ are shown in Table VI.

The diagonal terms are the multipliers of the values of a given cumulation. Thus the multipliers of the bottom entries of the columns of each of the three sets of cumulations of Table V are successively 1; -37, 46, 1369, -3323, 2116; etc.

This method is ideally adapted to the use of Hollerith cards. The information is punched on the cards to the number of places desired. The computational grouping is then accomplished by sorting. As an illustration we take the

TABLE VI

Multipliers when $a = -37$ and $k = 9$

	-37	-28	-19	-10	-1
46	1	46	2,116	97,336	4,477,456
55	-37	$-3,323$	$-222,969$	$-13,236,655$	
64	1,369	180,660	15,798,651		
73	$-50,653$	$-8,756,149$			
82	1,874,161				

TABLE VII

Hollerith Illustration

$x'' = x' - 4.5$	$C(f_{x.})$	$C(x_{f.})$	$C(y_{f.})$
210	2	422	1,387
200	6	1,242	4,154
190	10	2,017	6,901
180	21	4,035	14,733
170	48	8,727	33,675
160	110	18,923	76,565
150	250	40,466	173,491
140	458	70,392	316,073
130	719	105,431	492,774
120	900	127,990	613,763
110	980	137,203	666,088
100	990	139,199	678,299
90	999	139,199	678,299
80	1,000	139,288	678,896

records of the weights of 1000 students as reported by Carver [4] when measured to the nearest pound. The value of Σx is 139288 lbs. and that of Σx^2 is 19,692,450 $(\text{lb.})^2$ and we wish to obtain approximations to these values by grouping. If we let the grouping intervals be 80–89, 90–99, etc., with class marks $x' = 84.5$, 94.5, 104.5, etc., we would find by usual methods $\Sigma x' = 139{,}520$ lbs. and $\Sigma x'^2 = 19{,}760{,}430$ $(\text{lb})^2$. However, it is possible to wire in the three place number x, and to get from the same number of groups $\Sigma x = 139{,}288$ lbs. and $\Sigma xx' = 19{,}727{,}326$ $(\text{lb.})^2$. The unbiased values for method (1), (2), or (4) can be computed with the appropriate formulas of sections (3) and (4).

The Hollerith run is shown in Table VII where the first column indicates the smallest variate in the class rather than the class mark. The next columns show $C(f_{x'})$ and $C(x_{x'})$. The fourth column $C(y_{x'})$ is discussed in a later section. The values for method 3 and method 5 cannot be obtained so readily, since the quantities to be grouped are the x^2 and these do not appear on the card. However, it is possible to use a multiplying punch or to use a table of squares in the form of prepunched cards to get these values of x^2 on the cards. It might be preferable, in some cases, to do this work and then to use a coarser grouping than would be used otherwise.

6. Moments. The formulas (7), (8), (9), (10), give moment formulas if the proper values of F_x are assigned. We let $\nu_p = \dfrac{\Sigma x'^p}{N}$ and $\nu_{pq} = \dfrac{\Sigma x'^p x^q}{N}$ and have, in case $F_x = 1/N$ in (7) the usual formulas

$$
(17)\qquad
\begin{aligned}
E^{-1}[\mu_1] &= \nu_1 \\
E^{-1}[\mu_2] &= \nu_2 - R_2 \\
E^{-1}[\mu_3] &= \nu_3 - 3\nu_1 R_2 \\
E^{-1}[\mu_4] &= \nu_4 - 6\nu_2 R_2 + (6R_2^2 - R_4).
\end{aligned}
$$

If $F_x = x/N$ we have

$$
(18)\qquad
\begin{aligned}
E^{-1}[\mu_1] &= \mu_1 \\
E^{-1}[\mu_2] &= \nu_{11} \\
E^{-1}[\mu_3] &= \nu_{21} - R_2\mu_1 \\
E^{-1}[\mu_4] &= \nu_{31} - 3R_2\nu_{11}\,.
\end{aligned}
$$

While if $F_x = x^2/N$ we have

$$
(19)\qquad
\begin{aligned}
E^{-1}[\mu_2] &= \mu_2 \\
E^{-1}[\mu_3] &= \nu_{12} \\
E^{-1}[\mu_4] &= \nu_{22} - R_2\mu_2
\end{aligned}
$$

Similar formulas can be written for methods (4) and (5).

Previous to Carver's article in 1936 it was assumed that central moments could be used in place of moments in formulas (17) without introducing bias, but this article demonstrated that estimates obtained in this way are slightly biased. Thus

$$
(20)\qquad
\begin{aligned}
E(\bar\nu_2) &= E(\nu_2 - \nu_1^2) = E(\nu_2) - E(\nu_1^2) \\
&= \mu_2 + R_2 - \mu_2(\nu_1) = \mu_2 + R_2 - [\bar\mu_2(\nu_1) + \mu_1^2] \\
&= \bar\mu_2 + R_2 - \bar\mu_2(\nu_1) \text{ so that} \\
E^{-1}[\bar\mu_2] &= \bar\nu_2 - R_2 + \bar\mu_2(\nu_1)
\end{aligned}
$$

and so $\bar\nu_2 - R_2$ is a biased estimate of $\bar\mu_2$.

The general question of unbiased estimates of the central power sums and the central moments is one which has been studied for the conventional case by Pierce [3] and Craig [5] The more general discussion resulting from the introduction of the new methods is one which may well be deferred to a later paper. It is interesting to note in passing that the estimate of the variance obtained by substituting central moments for moments in method (2) is not biased since

$$E(\bar{\nu}_{11}) = E[\nu_{11} - \nu_1\mu_1] = \mu_2 - \mu_1^2 = \bar{\mu}_2 .$$

It is to be noted that the formulas previously used give correct results when the adjustments are defined to make the power sums and the moments rather than the central power sums and the central moments unbiased with respect to grouping. A sensible method of procedure in such a case is to make the correction on the power sum as soon as it is computed.

7. Product moments. Correlation. The introduction of additional variables opens up a variety of situations, since each of the variables may be grouped in different ways. Of these situations, one is immediately solved with the use of the formulas of section 3, and that is the case when one of the variables is not grouped. Let y be the ungrouped variable and let $F_x = y_{x'}$ be the sum of all the values of y having the same x grouped value, x'. This situation is frequently encountered when using Hollerith cards, as it is only necessary to wire in the whole variable y and take totals when the smallest value of x in the group is attained. Thus in Table VII, the value of $C(y_{x'})$ can be obtained simultaneously with the value of $C(f_{x'})$ and $C(x_{x'})$. Additional cumulation $C(z_{x'})$, $C(w_{x'})$, etc., could be obtained at the same time. It follows from Table VII that

$$\Sigma y = 678{,}896$$
$$E^{-1}\Sigma[xy] = \Sigma x'y = 94{,}929{,}322.$$

The actual value of Σxy is 94,774,336.

The general development of the theory of unbiased estimates of product moments is too extensive to be inserted here, but a brief outline might be indicated. We let the grouping errors be $\epsilon = x - x'$ and $\eta = y - y'$. Then the generalization of the lemma (4) is

$$\sum_x \sum_y x^a \epsilon^b y^c \eta^d F_x G_y = R_{b0} R_{0d} \sum_x x^a F_x \sum_y y^c G_y , \tag{21}$$

where R_{b0} is the bth central moment of the rectangular distribution consisting of ϵ's and R_{0d} is the dth central moment of the rectangular distribution consisting of η's. This is applied in turn to the terms of the expansion of

$$\Sigma x'^r y'^s F_x G_y .$$

For example

$$\begin{aligned} E[\Sigma x'y'F_xG_y] &= E\Sigma(x - \epsilon)(y - \eta)F_xG_y \\ &= \Sigma xyF_xG_y - R_{10}\Sigma yF_xG_y - R_{01}\Sigma xF_xG_y + R_{10}R_{01}\Sigma F_xG_y , \end{aligned} \tag{22}$$

and if $F_x = 1$ and $G_y = 1$, we have

$$E[\Sigma x'y'] = \Sigma xy \quad \text{so that} \quad E^{-1}[\Sigma xy] = \Sigma x'y'. \tag{23}$$

If we use the customary device of correcting the moments for bias, rather than the central moments or the ratio which is the correlation coefficient, we have the usual formula for correction of the correlation coefficient in which the numerator term is not corrected for bias, but the values in the denominator are corrected.

The use of method 2 gives $\Sigma x'y'$ and $\Sigma xy'$ as unbiased estimates of Σxy. It has been pointed out that these quantities $\Sigma x'y$ and $\Sigma xy'$ are readily obtained when the actual values of x and y are punched on Hollerith cards Each is in general a better estimate of Σxy than is $\Sigma x'y'$ since one of the values in the product, in each case, involves no approximation An average of these might be taken to obtain a better estimate of Σxy. If the values of $\Sigma x'$ and $\Sigma y'$ are also available, it is preferable to use the formula

$$E^{-1}[r] = \sqrt{\frac{A_{x'y} \cdot A_{xy'}}{A_{x'x} \cdot A_{yy'}}} \qquad \text{where } A_{xy} = N\Sigma xy - (\Sigma x)(\Sigma y). \tag{24}$$

The 1000 cards of weights and heights were used in this way with the digits grouped There resulted (dimensions omitted)

$$\begin{aligned} N &= 1{,}000 & & \\ \Sigma x &= 139{,}299 & \Sigma y &= 678{,}896 \\ \Sigma x' &= 139{,}520 & \Sigma y' &= 679{,}420 \\ \Sigma xx' &= 19{,}722{,}326 & \Sigma yx' &= 94{,}929{,}322 \\ \Sigma xy' &= 94{,}848{,}036 & \Sigma yy' &= 461{,}885{,}052 \end{aligned}$$

which gives $E^{-1}[r] = .4957$. The ungrouped 4 place value is .4952.

For use without Hollerith machines, this method indicates the recording of the values of $y_{x'}$ and $x_{y'}$ as well as $f_{x'y'}$ for each entry in the correlation chart.

The generalization of method 4 leads to

$$\Sigma xy = \Sigma(x' + \epsilon)(y' + \eta) = -\Sigma x'y' + \Sigma x'y + \Sigma xy' + \Sigma \epsilon\eta$$

so that we have

$$E^{-1}[\Sigma xy] = -\Sigma x'y' + \Sigma xy' + \Sigma x'y. \tag{25}$$

It is to be noted that the quantity $\Sigma x'y'$ is the unbiased estimate of Σxy resulting from the usual frequency distribution. This formula can be used with formula (11) of section 4 to obtain an estimate of the correlation coefficient.

The correlation chart application of method 4 demands the triple entry $f_{x'y'}$, $x_{y'}$, $y_{x'}$ for each of the squares of the correlation chart. From these values it is possible to compute all the entries needed to use method 4.

In general the values of $E[\Sigma x^{r} y^{t} F_x G_y]$ can be worked out with the repeated use of lemma 21. The reader who understands the developments of sections 3 and 4 should have little difficulty in writing out the formulas resulting here.

It should be pointed out that, in cases where the first and second order moments only are desired, it is frequently advisable to avoid grouping by using modern computing machines and, in this way, to eliminate the trouble and the errors caused by grouping [6].

8. Conclusion. There are additional points which might be considered, but they would take considerable space and the presentation is now sufficiently complete to enable one to obtain some perspective on the proper use of the new methods.

If precision is not needed, the use of the former grouping methods is advised. But if additional precision is needed, and if the results of a single grouping only are available, it is advised to use the newer methods. Method 2 is much more satisfactory than method 1 and, in many cases, will be sufficient, but, if additional precision is demanded, one can use method 3 or one of the compound methods

In general there are two kinds of groupings. One is a recorded grouping, and expresses the measures in terms of the units which are desired, while the second is a computational grouping which is introduced for the purpose of ease of computation. Now the recorded grouping, no matter whether obtained from discrete or continuous data, is necessarily discrete. Thus the weights, when measured, have to be recorded to the nearest pound, or the nearest tenth of a pound, or to the nearest hundredth of a pound, etc. The formulas to be applied to the results of computational grouping are the formulas for discrete variates. If in addition one wishes to correct continuous data for the recorded grouping, he may then apply the usual Sheppard's corrections for continuous data. However, it is advised to make the recording grouping sufficiently detailed so that the errors are slight. Thus one might record the values of heights to the nearest tenth of a pound, but use ten-pound intervals in making calculations. In this case the values when corrected for the computational grouping (to the nearest tenth of a pound) would presumably be sufficiently precise so that the additional grouping for recording would not be necessary. (In many cases the two grouping corrections are combined in a single grouping correction for continuous data)

It appears that it is not sufficiently satisfactory to continue to record the results of grouping in the usual form of a class mark (or class limits) and a frequency if the results are to be used by others. The table should include an additional column of $x_{x'}$ and preferably a column of $x_{x'}^2$, where the x' are the computational grouped values and the x are the measured values recorded to a considerable degree of accuracy. The arrangement takes little more space than the present frequency distribution, and it can be obtained from the recorded values with a reasonable amount of additional work. In the case of correlation

it is suggested that the present grouping of frequencies in the correlation chart be augmented with the values of $x_{y'}$ and $y_{x'}$ for each square. In this way it is possible for those who may use the distributions later to obtain much better estimates than would be possible from the frequency distributions as now recorded. This point certainly should be considered by all those who prepare tables for general use, and yet are forced by practical considerations to use some sort of grouping in reporting the results.

REFERENCES

[1] H. C. Carver, "The fundamental nature and proof of Sheppard's adjustments," *Annals of Math Stat*, Vol 7 (1936), pp 154-163

[2] J A Pierce, "A study of a universe of n finite populations with application to moment-function adjustments for grouped data," *Annals of Math. Stat.*, Vol. 11 (1940), pp 311-334.

[3] P. S. Dwyer, "The computation of moments with the use of cumulative totals," *Annals of Math Stat*, Vol. 9 (1938), pp. 288 303.

[4] H. C Carver, *Anthropometric Data*, Edwards Brothers, Ann Arbor, Mich., 1941.

[5] C. C. Craig, "Note on Sheppard's corrections," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 339 345

[6] P. S. Dwyer, "The calculation of correlation coefficients from ungrouped data with modern calculating machines," *Jour. of Am. Stat. Assn.*, Vol. 35 (1940), pp 671 673

[7] H L. Jones, "The use of grouped measurements," *Jour. of Am. Stat. Assn.*, Vol 36 (1941), pp. 525 529

ON THE CORRECT USE OF BAYES' FORMULA

By R. v. Mises
Harvard University

The problem that we try to solve by using Bayes' formula consists in making an inference from an observed statistical value upon the unknown value of a parameter, and in examining the chance of this inference being correct. One may call this the principle problem of practical statistics or the estimation problem, or, as the author put it in German (Rueckschluss-Wahrscheinlichkeit) problem of inference probability; at any rate we encounter this kind of problem in various forms in almost every branch of statistical investigation. It will be convenient to base the following discussion on a concrete question in quite specified form which will allow us to see clearer the points that are to be stressed in this paper.

1. The problem. In examining the quality of water supplies with respect to the number of bacterias of a certain kind they contain, a definite procedure is usually adopted. One takes $n = 5$ samples out of the water, each sample of exactly 10 ccm. Then by a certain biological test one finds out whether or not each sample contains at least one bacteria of the kind under consideration. The number x (zero to five) of positive tests is the observed value from which an inference is drawn upon the probability θ for a sample containing at least one bacteria. It is assumed that this θ is connected with the average number λ of bacterias per 10 ccm by

$$\theta = 1 - e^{-\lambda}; \qquad \theta = \theta_1 = 0.63 \qquad \text{for } \lambda = 1 \tag{1}$$

according to Poisson's law. A particular question which we want to answer is this: What is the chance of being right, if we conclude from the observed fact $x = 0$, (in other cases from $x = 1$) that θ lies between 0 and $\theta_1 = 0.63$ (or λ between 0 and 1)?

For a given θ the probability of getting x positive tests out of n tests is according to Bernoulli's formula

$$p(x \mid \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}. \tag{2}$$

The chance of having a θ-value between 0 and θ_1 when x positive tests are observed is according to Bayes' formula

$$P_x(\theta_1) = \frac{\int_0^{\theta_1} p(x \mid \theta)\, dP(\theta)}{\int_0^1 p(x \mid \theta)\, dP(\theta)} \tag{3}$$

where $P(\theta)$ is a distribution function, monotonically increasing from 0 to 1 and usually known as the *a priori probability*.

2. The apriori. The function $P(\theta)$ is generally considered as a troublemaker. As one uses to call P the a priori probability most people think that it has something to do with those absurd conceptions of non-empirical, a priori known probabilities that cannot be tested by any experiments etc. This cannot be strongly enough refuted. In our particular case the meaning of $P(\theta)$ is the following. Each probability statement refers, as we know, to a certain infinite sequence of experiments or trials which form a kollektiv. If we ask for the chance $P_x(\theta_1)$ of having a θ-value between zero and θ_1 when a certain x has been observed, we have in mind a sequence of trials each consisting of two steps, first, picking out one particular water supply, and then testing the number x of samples that contain bacillas. Among the first N trials of this kind we shall have N_1 cases where the θ-value for the water supply picked out lies between 0 and θ_1, then we shall have N_x cases where the number of positive tests is x, and finally in a number N_{1x} of cases both conditions will be fulfilled. The chance $P_x(\theta_1)$ we ask for is then by definition

$$P_x(\theta_1) = \lim_{N\to\infty} \frac{N_{1x}}{N_x}, \tag{4}$$

while the so-called a priori probability is

$$P(\theta_1) = \lim_{N\to\infty} \frac{N_1}{N}. \tag{5}$$

Later on we shall also use the probability

$$Q_x = \lim_{N\to\infty} \frac{N_x}{N}. \tag{6}$$

All these magnitudes are to the same extent empirical or non-empirical. They are "empirical," since we get approximate values for them out of a long sequence of experiments, and they may be considered as something super-empirical since the concepts of an infinite sequence and of a limit are used in the definition—as each theory must involve a certain amount of "idealization."

In order to avoid the above mentioned equivocation the author had suggested a long time ago[1] to call the probabilities corresponding to $P(\theta)$ and $P_x(\theta)$ respectively the *initial* and the *final* probability. Another expression which could be used in connection with the distribution function $P(\theta)$ is *overall distribution*, since it means the distribution of θ-values within the total mass of samples, not regarding what the values of x are in each case.

3. No randomness required. Now, the first remark we have to make is the following: In the Bayes' formula (3) the existence of a function $P(\theta)$ is presup-

[1] Cf. reference [2], p. 152.

posed, i.e. we assume that in the sequence of successive trials the frequency of those cases in which θ falls into a certain region has a definite limit. But nothing is assumed about this limit being independent of a place selection. The sequence of trials must fulfill the first condition of a kollektiv, with respect to θ but not the second; in other words *the randomness in the succession of θ-values is not required.* Thus we may say that θ is not supposed to be a chance variable in the usual sense of this term. Sometimes people are shocked by the idea that in Bayes' theory the individual cases are supposed to be picked out at random, and it is often considered as a superiority of the method of confidence intervals that here such assumption is avoided.

It is true that in the latter method even the existence of the frequency limit is not required,[2] but this does not seem to make any essential difference. The fact is that, if we want to make an inference upon the value of θ i.e. an assertion about the chance of θ falling into a certain interval, we have to assume that in the long run different θ-values may occur with certain frequencies.

It may be useful to have different expressions for the two cases where a frequency limit is or is not supposed to be independent of an arbitrary place selection. As we use the word probability in the first case it seems suitable to apply the word *chance* in the second. Thus, if $P(\theta)$ is the initial or the over all chance of θ we would say that $P_x(\theta_1)$ is the final chance of θ being smaller than or equal to θ_1 for a certain observed x-value. When $P(\theta)$ is supposed to be a probability, i.e. to fulfill the condition of randomness, then $P_x(\theta_1)$ will have this property too and has to be called probability.

4. Inequalities for the final chance $P_x(\theta)$. A much better founded objection against the practical application of Bayes' formula consists in saying that in most cases we have no sufficient information about the function $P(\theta)$. This undeniable fact leads often to an incorrect simplification of the formula by replacing in it $dP(\theta)$ by $d\theta$ which means an a priori probability of constant density. It is obvious that this is no solution: if you do not know what $P(\theta)$ is, to assume it equal to θ. On the other hand, if we accept Bayes' formula as correct (and there is no reason for not doing so) we learn that the value $P_x(\theta)$ we ask for *depends essentially on* $P(\theta)$, and is undetermined as far as $P(\theta)$ is undetermined. The only consequence in this situation is, first to use all information we can get about $P(\theta)$, and then to make the answer as vague or undetermined as the incompleteness of this information requires.

One way to do this consists in setting up inequalities for $P_x(\theta)$ based on certain inequalities for $P(\theta)$. A formula which turns out to be useful, at least in a well-known asymptotic problem is the following:

Let us consider the general case where θ stands for several variable parameters, and let A be the set of all possible values of θ. We are interested in the final probability P_C of a subset C of A given by

---

[2] Cf. reference [4], p. 201.

$$(6) \qquad P_C = \frac{\int_{(C)} p(x \mid \theta)\, dP(\theta)}{\int_{(A)} p(x \mid \theta)\, dP(\theta)},$$

where x is supposed to be known.

Let P'_C be the value of P_C under the assumption of a constant initial density and denote by P_B, P'_B the analogous values for a subset B which includes C so as to have

$$(7) \qquad C < B < A.$$

The quantities P'_B and P'_C depend only on the function $p(x \mid \theta)$ and the sets B and C while P_B and P_C change with $P(\theta)$.

If we assume that the initial density $p(\theta)$ has the limits

$$(8) \qquad \begin{aligned} m &\leqq p(\theta) \leqq M \quad \text{within } B \\ m' &\leqq p(\theta) \leqq M' \quad \text{within } A - B,\ (A \text{ minus } B) \end{aligned}$$

it can easily be shown that

$$(9) \qquad \frac{m}{M} P'_B + \frac{m'}{M}(1 - P'_B) \leqq \frac{P'_C}{P_C} \leqq \frac{M}{m} P'_B + \frac{M'}{m}(1 - P'_B).$$

We may consider the following application of these inequalities.

If we are concerned with a case where a great number n of trials is involved, the function $p(x \mid \theta)$—which determines the P' values—shows an increasing concentration at a certain point of the set A. In other words, for large n we have a subset B more and more reducing to one single point for which P'_B is as near to 1 as we want. If we then assume that the density $p(\theta)$ is continuous and bounded, the difference between m and M tends to zero, and if m is supposed to have a positive lower bound, both the first and the last expression in (9) tend to unit or P_C approaches P'_C. This is a generalized form of the statement which the author proved for the first time in 1919,[3] that in the original Bayes' problem where we are concerned with n repetitive observations of an alternative, *the final probability becomes more and more independent of the initial probability* $P(\theta)$ as the number n of observations involved increases.

5. Using previous experience. The inequalities (9) may be of use in many cases. But to be sure, in general, they are not the basis upon which practical estimation judgments rest. Everybody acquainted with the conditions of testing water supplies takes it for granted that the outcome $x = 0$ (no positive test) supplies a sufficient reason for the statement $\theta \leqq \theta_1 = 0.63$ (less than one

[3] Cf. reference [1], p. 81

bacteria per 10 cc). But, if nothing were known about the initial distribution $P(\theta)$, we could assume $P(\theta)$ in the form

$$P(\theta) = \theta^m, \qquad p(\theta) = m\theta^{m-1} \qquad \text{for } 0 \leqq \theta \leqq 1,$$

with a large value of m. With $n = 5$, $x = 0$ equations (2) and (3) give $P_0(\theta_1) = 0.50$ for $m = 10$, and $P_0(\theta_1) = 0.88$ if m is 5. These values are much too low to justify any recommendation of a water supply for which x was found to be zero. Thus we have to ask: What is the *real source of the confidence* we put in the inference from $x = 0$ upon $\theta \leqq \theta_1$?

There is no doubt, that this confidence is based on previous experience We know that the water supplies subjected to the routine test in the past formed a class of rather clean than dirty water and we rely that a new sample will belong to the same class. The author was given the following information about the results under the jurisdiction of Massachusetts during the last decade. Out of a total of $N = 3420$ examinations there were found

3086 cases with $x = 0$ (no positive test)

279 cases with $x = 1$ (one positive test)

32 cases with $x = 2$

15 cases with $x = 3$

5 cases with $x = 4$

3 cases with $x = 5$

The overwhelming majority of cases with $x = 0$ is evident. The question is only how we can use these statistics of past experiments for obtaining a numerical inference upon the value of $P_x(\theta)$.

If the initial distribution $P(\theta)$ were known, we could find the probability Q_x of getting x positive tests out of n:

$$Q_x = \int_0^1 p(x \mid \theta)\, dP(\theta) = \binom{n}{x} \int_0^1 \theta^x (1 - \theta)^{n-x}\, dP(\theta). \tag{10}$$

Using the numbers N_1, N_x, N_{1x} introduced in section 2 the probability $Q(x)$ is defined by equation (6).

If the number N of past examinations is considered as sufficiently large, we can take the ratios 3086/3420, 279/3420 etc. as approximate values for Q_0, Q_1 etc. Now, according to the well-known identities

$$\frac{1}{n} \sum_{x=0}^{n} x \binom{n}{x} \theta^x (1 - \theta)^{n-x} = \theta, \tag{11}$$

$$\frac{1}{n(n-1)} \sum_{x=0}^{n} x(x-1) \binom{n}{x} \theta^x (1 - \theta)^{n-x} = \theta^2, \tag{12}$$

and using (10) we can derive from the values $Q_0, Q_1, \cdots, Q_n$ *the first and second moments* of the distribution function $P(\theta)$.

$$
\begin{aligned}
M_1 &= \int_0^1 \theta\, dP(\theta) = \frac{1}{n} \sum_{x=0}^{n} Q_x \\
M_2 &= \int_0^1 \theta^2\, dP(\theta) = \frac{1}{n(n-1)} \sum_{x=0}^{n} x(x-1) Q_x .
\end{aligned} \tag{13}
$$

If we introduce here the above mentioned empirical ratios for Q_x we find the approximate values for the first and second moments of $P(\theta)$:

$$M_1 = 0.02474 \qquad M_2 = 0.00401. \tag{13'}$$

6. Determination of a distribution function by its first moments. In an earlier paper the author showed [3] how the exact upper and lower bounds for a distribution function $P(\theta)$ can be found, if the expected values of two functions $f(\theta)$ and $g(\theta)$ are known. The only condition was that the curve represented in a Cartesian coordinate system by $x = f(\theta)$, $y = g(\theta)$ is convex. Let us take

$$
\begin{aligned}
& f(\theta) = g(\theta) = 0 && \text{for } \theta < 0 \\
& f(\theta) = \theta, \qquad g(\theta) = \theta^2 && \text{for } 0 \leqq \theta \leqq 1 \\
& f(\theta) = g(\theta) = 1 && \text{for } \theta > 1.
\end{aligned} \tag{14}
$$

In this case the condition is fulfilled and the expected values of $f(\theta)$ and $g(\theta)$ are the moments M_1, M_2, respectively. The results obtained in the paper quoted above take the following form:

First, we have to derive from the given values M_1 and M_2 two points θ' and θ'' of the internal $0 \leqq \theta \leqq 1$

$$\theta' = \frac{M_1 - M_2}{1 - M_1}, \qquad \theta'' = \frac{M_2}{M_1}. \tag{15}$$

Then the limits for $P(\theta)$ are:

$$
\begin{aligned}
& 0 \leqq P(\theta) \leqq \frac{M_2 - M_1^2}{M_2 - 2M_1\theta + \theta^2} && \text{for } 0 \leqq \theta \leqq \theta' \\
& 1 - M_1 - \frac{M_1 - M_2}{\theta} \leqq P(\theta) \leqq 1 - \frac{M_1\theta - M_2}{\theta - 1} && \text{for } \theta' \leqq \theta \leqq \theta'' \\
& \frac{(M_1 - \theta)^2}{M_2 - 2M_1\theta + \theta^2} \leqq P(\theta) \leqq 1 && \text{for } \theta'' \leqq \theta \leqq 1.
\end{aligned} \tag{16}
$$

In our case we find $\theta' = 0.0213$, $\theta'' = 0.1619$ and the point $\theta_1 = 0.6321$ falls into the third interval θ'', 1. The lines $O\ A\ B\ C$ and $O\ D\ E\ F\ G$ in Fig. 1 show (slightly distorted) the lower and upper bounds for $P(\theta)$.

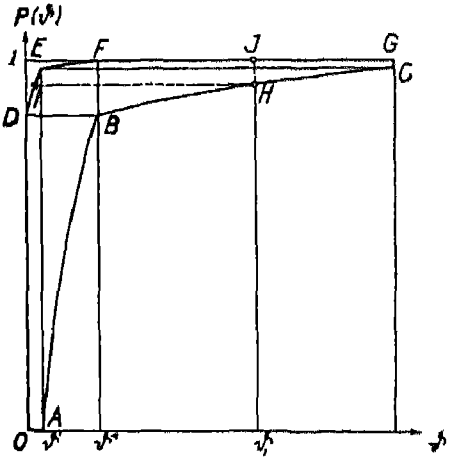


Fig. 1

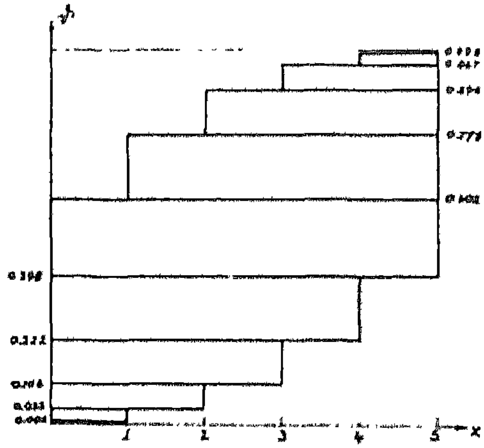


Fig. 2

Fig. 1. The limits of the overall distribution function
Fig. 2. The 99% region in the methods of confidence intervals

7. Application to Bayes' formula. The inequalities (16) enable us to find in a simple way a lower bound for the end probability $P_x(\theta_1)$ defined by (2) and (3) in the case $x = 0$. Let us denote by A the numerator in (3) and by B the supplementary integral

$$B = \int_{\theta_1}^{1} p(x \mid \theta)\, dP(\theta), \tag{17}$$

so as to have $A + B$ for the denominator in (3). If the subscripts min and max denote a lower and upper bound respectively we can write

$$P_x(\theta_1) = \frac{A}{A + B} \geqq \frac{A_{\min}}{A_{\min} + B_{\max}}. \tag{18}$$

Now, taking $x = 0$ we find by product integration

$$A = P(\theta_1)(1 - \theta_1)^n + n \int_0^{\theta_1} P(\theta)(1 - \theta)^{n-1}\, d\theta. \tag{19}$$

Therefore, $A_{\min}$ is found when we introduce in this expression the lower lim for $P(\theta)$ as given in (16). If we do this and use the values for M_1 and M_2 according to (13′), numerical computation leads to $A_{\min} = 0.712$.
In the same way we obtain B in the form

$$B = -P(\theta_1)(1 - \theta_1)^n + n \int_{\theta_1}^{1} P(\theta)(1 - \theta)^{n-1}\, d\theta. \tag{20}$$

The upper bound $B_{\max}$ is reached, if we introduce in the integral $P(\theta) = 1$ and in the first term the minimum value for $P(\theta_1)$ following from (16). The second

term becomes thus equal to $(1 - \theta_1)^n$ and the numerical result is $B_{max} = 0.0000607$. Therefore the inequality (18) supplies

$$P_0(\theta_1) \geqq \frac{0.712}{0.71206} = 0.99915. \tag{18'}$$

The final outcome secured in this way can be formulated as follows: *If we assume that in continuing the experiments the distribution of test results will be about the same as it has been in the past* 3420 *cases, we have a chance of more than* 99.9% *of being right, when we state in each case of no positive test that the density of bacterias is less than* 1 *per* 10 *ccm.*

The high value of 99 9% for $P(\theta_1)$ is of course strictly bound to the assumption that the entire mass of water supplies to be tested is homogeneous and sufficiently characterized by the distribution of test results found in the past. If e.g we had to assume that the six possible values for x (0 to 5) in the long run appear with equal frequencies so as to have $Q_0 = Q_1 = \cdots Q_5 = \frac{1}{6}$, the same method would give $M_1 = \frac{1}{2}$, $M_2 = \frac{1}{3}$, then $\theta' = \frac{1}{3}$, $\theta'' = \frac{2}{3}$, and the final result would be $P_0(\theta_1) \geqq 0.73$. The assumption of a constant initial density $P(\theta) = \theta$ would give $P_0(\theta_1) = P_0'(\theta_1) = 0.9975$, a little less than the value found above in (18').

8. The case $x = 1$. The results are less favorable in the case of one positive test, $x = 1$. Here we have

$$p(1 \mid \theta) = n\theta(1 - \theta)^{n-1} = 5\theta(1 - \theta)^4, \tag{21}$$

and the derivative of p is first positive, then negative. We can conclude from Fig. 1 that the minimum value for A and the maximum for B will be reached when the distribution function $P(\theta)$ is represented by the line $O\ D\ I\ H\ J\ G$ where $I\ H$ is horizontal and H the point on $B\ C$ with abscissa θ_1. The abscissa θ_0 of I is determined by the equation

$$\frac{M_2 - M_1^2}{M_2 - 2M_1\theta_0 + \theta_0^2} = \frac{(M_1 - \theta_1)^2}{M_2 - 2M_1\theta_1 + \theta_1^2}, \tag{22}$$

which supplies $\theta_0 = 0.0190$. We then have

$$A_{min} = \int_0^{\theta_0} p(1 \mid \theta)\, dP(\theta), \tag{23}$$

with the value $p(1 \mid \theta)$ from (21) and with

$$P(\theta) = \frac{M_2 - M_1^2}{M_2 - 2M_1\theta + \theta^2}$$

according to (16). On the other hand B_{max} is found, as in the former case, to be

$$B_{max} = p(1 \mid \theta_1)[1 - P(\theta_1)], \tag{24}$$

where we have to take for $P(\theta_1)$ its minimum value according to (16). The numerical computation yields $A_{min} = 0.0062$ and $B_{max} = 0.00052$ so as to give

$$P(\theta_1) \geqq \frac{62}{67.2} = 0.92 .$$

The result is that under the assumption above mentioned *we have more than 92% chance of being right*, if we predict each time one out of five tests has been positive that the density of bacilli is less than 1 per 10 ccm. The chance computed under the assumption of a uniform initial distribution $P(\theta) = \theta$ would be 0.97.

9. The method of confidence intervals. One may ask what kind of answer to our questions can be deduced from the principle of confidence intervals. This method has undeniably to its credit that no use is made here of the initial distribution $P(\theta)$ and that, therefore, all its statements are completely independent of what is assumed about $P(\theta)$.

In order to apply this method[4] we have to select for a given degree of confidence, say $\alpha = 0.99$, a region of acceptance, i.e. an area in the two dimensional x, θ plane limited by two lines $x_1(\theta)$ and $x_2(\theta)$ so as to have for each θ

$$\text{Prob}\,\{x_1(\theta) \leqq x \leqq x_2(\theta)\} = \alpha. \tag{25}$$

The region is, of course, not uniquely determined by (25). In our case, however, one will generally agree that the best way to determine the region consists in assuming for $x_1(\theta)$ and $x_2(\theta)$ two step lines with steps at the integer values $x = 0, 1, 2, \cdots$ as indicated in Fig. 2. Then the formula (2) for $p(x \mid \theta)$ combined with (25) supplies the abscissae of the steps, if x is given. If we transform the limits for θ into limits for λ using equation (1), the final outcome reads as follows:

Whatever the initial distribution $P(\theta)$ may be, we have a chance of 99% of being right, if we predict:

each time $x = 0$ *is observed that* λ *lies between* 0 *and* 0.92,
each time $x = 1$ *is observed that* λ *lies between* 0.002 *and* 1.51,
each time $x = 2$ *is observed that* λ *lies between* 0.036 *and* 2.24,
each time $x = 3$ *is observed that* λ *lies between* 0.112 *and* 3.41,
each time $x = 4$ *is observed that* λ *lies between* 0.25 *and* 8.48,
each time $x = 5$ *is observed that* λ *lies between* 0.51 *and* ∞.

It is true that in this way we obtain a result independent of any assumption on $P(\theta)$. But it is essential that the chance of $\alpha = 99\%$ holds only *for the six joint statements as a whole*. This means it may happen that for instance the first assertion (that λ is smaller than 0.92 in the case $x = 0$) is correct but very seldom or even never, while other assertions (e.g. those for $x = 4$ and 5) have

[4] Cf. reference [5] and reference [4], p. 203.

a much greater chance than 99% of being correct. Whether this happens or not depends on the initial distribution $P(\theta)$. As long as we know nothing about $P(\theta)$ we are not in the position to conclude, by using the method of confidence intervals, that the particular statement "$\lambda \leqq 0.92$ if $x = 0$" has a chance of 99% or even any chance at all of being correct. On the other hand, when $x = 0$ has been observed we are in no way interested in consequences that may be drawn in the case $x = 4$ or $x = 5$ or in a set of statements that includes the cases $x = 4$ and $x = 5$. The only practical question that is relevant to the purpose for which the tests are made is this. *What can we conclude from the fact that in a certain instance $x = 0$ has been observed* (or in another instance $x = 1$)? It seems that the method of confidence intervals, discarding any consideration of the initial distribution, can supply no contribution towards the answering *this particular question.*

REFERENCES

[1] R. v. Mises, "Fundamentalsaetze der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol. 4 (1919), pp. 1–97

[2] R. v. Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik*, Leipzig und Wien, 1931

[3] R. v. Mises, "The limits of a distribution function if two expected values are given," *Annals of Math. Stat.*, Vol. 10 (1939), pp 99–104

[4] R. v. Mises, "On the foundation of probability and statistics," *Annals of Math. Stat* Vol. 12 (1941), pp. 191–205.

[5] J Neyman, *Roy Stat Soc. Jour.*, Vol 97 (1934), pp 590–592

AN ITERATIVE METHOD OF ADJUSTING SAMPLE FREQUENCY TABLES WHEN EXPECTED MARGINAL TOTALS ARE KNOWN

By Frederick F. Stephan

Cornell University and U. S. Census Bureau

1. Introduction. In a previous paper by W. Edwards Deming and the author [1] the method of least squares was applied to the adjustment of sample frequency tables for which the expected values of the marginal totals are known. From observations on a sample the frequencies n_{ij} for the cell in the ith row and jth column of a two dimensional table and the r row and s column totals, $n_{i.}$ and $n_{.j}$, are obtained. These frequencies are subject to the errors of random sampling and it is desired to adjust them so that the row and column totals will agree with their expected values, $m_{i.}$ and $m_{.j}$, which are known. The adjustment involves the solution of the $r + s - 1$ normal equations

$$(1)\qquad \begin{aligned} n_{i.}\lambda_{i.} + \sum_j n_{ij}\lambda_{.j} &= m_{i.} - n_{i.} \qquad i = 1, 2, \cdots, r \\ \sum_i n_{ij}\lambda_{i.} + n_{.j}\lambda_{.j} &= m_{.j} - n_{.j} \qquad j = 1, 2, \cdots, s - 1 \end{aligned}$$

where the λ are Lagrange multipliers from which are calculated the adjusted frequencies

$$(2)\qquad m_{ij} = n_{ij}(1 + \lambda_{i.} + \lambda_{.j}).$$

Similar equations arise in the three dimensional case.

A method of iterative proportions was presented for effecting the adjustments more conveniently than by solving the normal and condition equations, and it was stated that "the final results coincide with the least squares solution." This statement is incorrect, for although the adjusted values satisfy the condition equations, they do not satisfy the normal equations and hence they provide only an approximation to the solution. The method of iterative proportions has several interesting characteristics that will be discussed in a later section. This paper now presents a method that converges to the values given by the least squares adjustment and is self correcting. It can be used with any set of data and weights for which a least squares solution exists. The two-dimensional case will be considered first.

2. The two-dimensional case; expected row and column totals known. Assume that a sample of n items is drawn at random and cross-classified in a table of r rows and s columns. As in the previous paper, let n_{ij} be the frequency in the ith row and jth column of the two-way frequency distribution. Indicate summation by substituting a dot for the letter over which the summation is to be performed. Then $n_{i.}$ and $n_{.j}$ are the marginal totals for the ith row and jth column respectively. Let $m_{i.}$ and $m_{.j}$ be the expected values of these

marginal totals calculated from other information or from theoretical considerations, and c_{ij} a set of constants known or estimated to be proportional to the reciprocals of the weights of the n_{ij}, i e proportional to their error variances. Since the weights are positive, the c_{ij} are non-negative and finite. It is assumed that the set of weights is such that for the given data an adjustment exists.

The least squares adjusted frequencies m_{ij} can be computed from the given numbers c_{ij}, n_{ij}, $m_{i.}$, and $m_{.j}$ by a series of approximate adjustments in a manner now to be explained Let $m_{ij}^{(p)}$ be the pth approximation to m_{ij}. In conformity with this notation $m_{ij}^{(0)} = n_{ij}$. Let

$$(3) \quad d_{ij}^{(p)} = m_{ij} - m_{ij}^{(p)}, \qquad d_{i.}^{(p)} = m_{i.} - m_{i.}^{(p)}, \qquad d_{.j}^{(p)} = m_{.j} - m_{.j}^{(p)},$$

be corrections that must be added to the $m^{(p)}$ to produce the least squares adjusted frequencies. As $d \to 0$, $m^{(p)} \to m$. Let $\lambda_{i.}^{(p)}$ and $\lambda_{.j}^{(p)}$ be constants determined arbitrarily between the limits set by equations (5) to (7). Any one λ may be fixed arbitrarily and kept constant through successive approximations. Note that $\lambda_{i.}^{(0)} = \lambda_{.j}^{(0)} = 0$ and that, if at every step we set $\lambda_{.s}^{(p)} = 0$, the $\lambda^{(p)}$ are approximations to the Lagrange multipliers in the normal equations. After p steps in the iterative process the approximate adjusted frequencies will be

$$(4) \qquad m_{ij}^{(p)} = n_{ij} + c_{ij}(\lambda_{i.}^{(p)} + \lambda_{.j}^{(p)}).$$

The following conditions, derived from (19), (23), and (24), are sufficient to make the successive approximations converge to the least squares adjusted frequencies:

$$(5) \qquad \begin{aligned} \lambda_{i.}^{(p)} &= \lambda_{i.}^{(p-1)} + \theta_{i.}^{(p)}\, d_{i.}^{(p-1)}/c_{i.}, \\ \lambda_{.j}^{(p)} &= \lambda_{.j}^{(p-1)} + \theta_{.j}^{(p)}\, d_{.j}^{(p-1)}/c_{.j}, \end{aligned}$$

$$(6) \qquad 0 \leq \theta_{i.}^{(p)}, \qquad 0 \leq \theta_{.j}^{(p)}, \qquad \theta_{i.}^{(p)} + \theta_{.j}^{(p)} \leq 2,$$

and, for at least one pair ij,

$$(7) \qquad \theta_{i.}^{(p)}(d_{i.}^{(p-1)})^2 + \theta_{.j}^{(p)}(d_{.j}^{(p-1)})^2 > 0; \qquad \theta_{i.}^{(p)} + \theta_{.j}^{(p)} < 2.$$

The θ's are introduced because in actual computations the successive approximations $\lambda^{(p)}$ can only be calculated to a limited number of digits and because the adjustment may progress more rapidly if the computer is permitted to use his judgment in determining the approximations as he observes the course of previous approximations.

The process of adjustment is continued until the $d_{i.}^{(p)}$ and $d_{.j}^{(p)}$ become sufficiently small to provide the desired degree of agreement between the adjusted and expected row and column totals.

3. Example. The following example shows the steps in the adjustment for a table of 3 rows and 4 columns with $\theta_{i.}^{(p)} \doteq \theta_{.j}^{(p)} \doteq 1$:

ij	n_{ij}	m_{ij}	$d_{ij}^{(0)}$	c_{ij}	$\lambda^{(0)}$	$m_{ij}^{(1)}$	$d_{ij}^{(1)}$	$d_{ij}^{(1)}/c_{ij}$	$\lambda^{(1)}$	$m_{ij}^{(2)}$	$d_{ij}^{(2)}$	m_{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
11	783	—	—	75	—	777.5				772.[illegible]		771
12	7426	—	—	455	—	7505.6				7496.4		7497
13	4709	—	—	358	—	4712.6				4709.6		4711
14	2145	—		176	—	2055.8				2051.3		2049
21	517	—	—	52		528.9				529.5		529
22	928	—	—	95	—	973.1				978.4		979
23	622	—	—	56	—	639.5				643.1		644
24	703	—	—	70	—	688.7				691.9		692
31	207	—	—	19	—	200.3				201.1		201
32	373	—	—	38	—	369.1				372.3		373
33	337	—	—	31	—	328.7				331.7		332
34	425	—	—	39	—	391.5				397.5		397
.1	1507	1501	−6	146	−.041	1506.7	−5.7	−.0390	−.0800	1503.5	−2.5	1501
.2	8727	8849	+122	588	+.208	8848.1	+0.9	+.0015	+.2095	8846.9	+2.1	8849
.3	5668	5687	+19	415	+.043	5680.8	+6.2	+.0139	+.0569	5684.1	+2.6	5687
.4	3273	3138	−135	285	−.474	3139.0	−1.0	−.0035	−.4775	3140.7	−2.7	3138
1.	15063	15028	−35	1064	−.033	15051.5	−23.5	−.0221	−.0551	15030.1	−2.1	15028
2.	2770	2844	+74	273	+.27	2830.5	+13.5	+.0495	+.3195	2842.8	+1.2	2844
3.	1342	1303	−39	127	−.31	1292.6	+10.4	+.0819	−.2281	1302.6	+0.4	1303
	19175	19175	0	1464	—	19174.9	+0.1	—	—	19175.5	−0.5	19175

Columns (1), (2) and (4) are given. Columns (3) and (6) to (11) are calculated in succession using equations (3), (4), and (5). It is not necessary in practice to record the θ's or even determine their values since the $\lambda^{(p)}$ may be determined directly at convenient values approximately equal to their corresponding $\lambda_{i.}^{(p-1)} + d_{i.}^{(p-1)}/c_{i.}$ and $\lambda_{.j}^{(p-1)} + d_{.j}^{(p-1)}/c_{.j}$. The final adjusted frequencies given in column (12) are derived by another repetition of the adjustment process but the amounts involved are so small that they can be calculated mentally and the n_{ij} rounded at the same time.

4. **Computing procedure.** The computing procedure may be set up in any of a number of ways to meet the preferences of the computer and the characteristics of the problem. Ordinarily it is desirable to make every number positive and the procedure as nearly routine as possible.

For two-dimensional adjustments the following procedure of computing alternately by columns and by rows is convenient:

(a) Set up a table of the c_{ij} in r rows and s columns. Enter the $c_{i.}$ in the $s + 1$ column, the $c_{.j}$ in the $r + 1$ row, and $c_{..} = \sum_i c_{i.} = \sum_j c_{.j}$ in the common cell.

(b) Calculate the quantities $A_{i.} \doteq (d_{i.}^{(0)}/c_{i.}) + a$ and $A_{.j} \doteq (d_{.j}^{(0)}/c_{.j}) + a$ and enter them in the $s + 2$ column and $r + 2$ row. The constant a is selected at some value that will make all quantities in the computations positive and may be any convenient integer greater than $2 \max | d_{i.}^{(0)}/c_{i.} |$ or $2 \max | d_{.j}^{(0)}/c_{.j} |$.

(c) Calculate the factors $\mu_{i.}^{(1)}$ approximately equal to the $A_{i.} - \frac{1}{2}a$ and enter each on its corresponding row in the $s + 3$ column. Throughout the computations the μ^p are merely $\lambda_{ij} + \frac{1}{2}a$.

(d) Take column j and multiply each c_{ij} by its corresponding $\mu_{i.}^{(1)}$ accumulating the products in the calculating machine. Divide the sum of products by $c_{.j}$, subtract the quotient from $A_{.j}$, and record the difference $\mu_{.j}^{(2)}$ in the jth column on the $r + 3$ row. Repeat for each of the other columns.

(e) Take row i and multiply each c_{ij} by its corresponding $\mu_{.j}^{(2)}$ accumulating the products in the calculating machine. Divide the sum of products by $c_{i.}$, subtract the quotient from $A_{i.}$, and record the difference $\mu_{i.}^{(3)}$ on the ith row in the $s + 4$ column bordering the table on the right. Repeat for each of the other rows.

(f) Repeat steps (d) and (e) alternately until a satisfactory degree of stability is reached in the $\mu_{i.}$ and $\mu_{.j}$. Then compute each adjusted frequency as follows:

$$m_{ij}^{(p)} = c_{ij}(\mu_{i.}^{(p)} + \mu_{.j}^{(p)} - a) + n_{ij}, \tag{8}$$

taking either $\mu_{i.}^{(p)} = \mu_{i.}^{(p-1)}$ or $\mu_{.j}^{(p)} = \mu_{.j}^{(p-1)}$ as the case may be.

(g) The computations may be checked at any step by computing

$$\sum_j \mu_{.j}^{(p)} c_{.j} = \sum_j A_{.j} c_{.j} - \sum_i \mu_{i.}^{(p-1)} c_{i.} = ac_{..} - \sum_i \mu_{i.}^{(p-1)} c_{i.}, \tag{9}$$

or

$$\sum_i \mu_{i.}^{(p)} c_{i.} = \sum_i A_{i.} c_{i.} - \sum_j \mu_{.j}^{(p-1)} c_{.j} = ac_{..} - \sum_j \mu_{.j}^{(p-1)} c_{.j}. \tag{10}$$

(h) At any step a constant may be added to all the $\mu_{i.}^{(p)}$ and subtracted from all the $\mu_{.j}^{(p)}$; this may be necessary to keep the μ's all positive. It has no effect on the value of a to be used in (8).

(i) If it is desired to "inflate" the adjusted frequencies $(\sum_{i,j} m_{ij} \neq \sum_{i,j} n_{ij})$ first multiply each n_{ij}, $n_{i.}$, and $n_{.j}$ by the factor $\sum_{i,j} m_{ij}/\sum_{i,j} n_{ij}$ and then proceed as above using the products in place of their corresponding n_{ij}, $n_{i.}$ and $n_{.j}$.

(j) If before the iterative process has reached an acceptable adjustment it is desired to force a satisfaction of the condition equations, compute:

$$m_{ij}^{(p+1)} = c_{ij}(\mu_{i.}^{(p)} + \mu_{.j}^{(p)} - a) + n_{ij} + (d_{i.}^{(p)} c_{.j} + d_{.j}^{(p)} c_{i.})/c_{..}, \tag{11}$$

in which either the $d_{i.}^{(p)}$ or the $d_{.j}^{(p)}$ are all zero.

5. **Adjustments in three dimensions.** If the sample is cross-tabulated in a three-way frequency distribution, there are two cases that are not reducible to

two-way distributions. These are designated Case III and Case VII in the earlier paper [1]. The adjustment equations are, respectively,

$$
(12) \qquad
\begin{aligned}
m_{ijk}^{(p)} &= n_{ijk} + c_{ijk}(\lambda_{i..}^{(p)} + \lambda_{.j.}^{(p)} + \lambda_{..k}^{(p)}) \\
m_{ijk}^{(p)} &= n_{ijk} + c_{ijk}(\lambda_{ij.}^{(p)} + \lambda_{i.k}^{(p)} + \lambda_{.jk}^{(p)}),
\end{aligned}
$$

subject to conditions on the choice of the λ corresponding to equations (5), (6), and (7). For Case III, the conditions are that

$$
(13) \qquad 0 \leq \theta_{i..}^{(p)}, \quad 0 \leq \theta_{.j.}^{(p)}, \quad 0 \leq \theta_{..k}^{(p)}, \quad \theta_{i..}^{(p)} + \theta_{.j.}^{(p)} + \theta_{..k}^{(p)} \leq 2,
$$

and for at least one triple ijk, $\theta_{i..}^{(p)}(d_{i..}^{(p-1)})^2 + \theta_{.j.}^{(p)}(d_{.j.}^{(p-1)})^2 + \theta_{..k}^{(p)}(d_{..k}^{(p-1)})^2 > 0$ and $\theta_{i..}^{(p)} + \theta_{.j.}^{(p)} + \theta_{..k}^{(p)} < 2$. Similar conditions apply to Case VII.

The computing procedure described in Section 4 can be extended readily to the three-dimensional case. For example, in Case VII calculate $\mu_{ij.}^{(1)}$ approximately equal to $(d_{ij.}^{(0)}/c_{ij.}) + \frac{1}{3}a$ and $\mu_{i.k}^{(1)}$ approximately equal to $(d_{i.k}^{(0)}/c_{i.k}) + \frac{1}{3}a$. Then multiply each c_{ijk} in the column jk by its corresponding $(\mu_{ij.}^{(1)} + \mu_{i.k}^{(1)})$ accumulating the products in the calculating machine. Divide the sum of the products by $c_{.jk}$ and subtract the quotient from $(d_{.jk}^{(0)}/c_{.jk}) + a$. Record the difference as $\mu_{.jk}^{(2)}$ and repeat the process for every other jk column. Take $\mu_{ij.}^{(2)} = \mu_{ij.}^{(1)}$ and repeat for each ik column to obtain $\mu_{i.k}^{(3)}$; then take $\mu_{.jk}^{(3)} = \mu_{.jk}^{(2)}$ and repeat for each ij column to obtain $\mu_{ij.}^{(4)}$ and so on. The final adjusted frequencies are

$$
(14) \qquad m_{ijk}^{(p)} = n_{ijk} + c_{ijk}(\mu_{ij.}^{(p)} + \mu_{i.k}^{(p)} + \mu_{.jk}^{(p)} - a).
$$

6. **The general case.** The iterative method can be extended readily to more than three dimensions and to various systems of condition equations. A simple general notation may now be introduced. Let the cells be numbered in any order from 1 to t and for the ith cell let n_i be the value given by the sample, c_i a finite positive constant known or estimated to be inversely proportional to the weight of n_i, m_i the least squares adjusted value to be determined, $m_i^{(p)}$ the pth approximation to m_i, $d_i^{(p)} = m_i - m_i^{(p)}$, and $m_i^{(0)} = n_i$. Assume that the values m_σ of certain linear combinations of the m_i are given, i.e. there is a system of consistent linear equations of condition numbered in any order, the σth equation being

$$
(15) \qquad \sum_i b_{i\sigma} m_i = m_\sigma, \qquad \sum_i b_{i\sigma}^2 > 0,
$$

$b_{i\sigma}$ and m_σ being known a priori. The corresponding linear combinations of the n_i and $d_i^{(p)}$ define

$$
(16) \qquad n_\sigma = \sum_i b_{i\sigma} n_i, \qquad d_\sigma^{(p)} = \sum_i b_{i\sigma} d_i^{(p)}.
$$

Let

$$
(17) \qquad c_\sigma = \sum_i b_{i\sigma}^2 c_i .
$$

The pth approximation to m_i is

$$m_i^{(p)} = n_i + c_i \sum_\sigma b_{i\sigma}\lambda_\sigma^{(p)}, \tag{18}$$

where

$$\lambda_\sigma^{(p)} = \lambda_\sigma^{(p-1)} + \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma, \qquad \lambda_\sigma^{(0)} = 0, \tag{19}$$

the $\theta_\sigma^{(p)}$, and therefore the $\lambda_\sigma^{(p)}$, being arbitrary for a finite number of steps, say p', but determined thereafter so that

$$2 \sum_\sigma \theta_\sigma^{(p)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma}\theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma)^2 \geq (d_\tau^{(p-1)})^2/(c_\tau H), \tag{20}$$

τ being a value of σ, chosen at the pth step, for which $(d_\sigma^{(p-1)})^2/c_\sigma$ is a maximum and H a finite number greater than 1 fixed prior to the first step as large as one will. That this condition can be satisfied may be shown by putting $\theta_\tau^{(p)} = 1$ and $\theta_\sigma^{(p)} = 0$ $(\sigma \neq \tau)$.

A weighted average of several of the possible selections of $\theta_\sigma^{(p)}$ satisfying (20) will also satisfy (20), positive "weights" being assumed Let k added to the superscript represent the kth such selection and let $\alpha^{(p,k)} > 0$ be a constant for "weighting" the kth selection in the weighted average which may be chosen arbitrarily except that $\sum_k \alpha^{(p,k)} = 1$. Then, if the kth selection of $\theta_\sigma^{(p)}$ is represented by $\theta_\sigma^{(p,k)}$, the weighted averages are $\theta_\sigma^{(p,0)} = \sum_k \alpha^{(p,k)}\theta_\sigma^{(p,k)}$. Substitute them in the left-hand side of (20),

$$\begin{aligned}
&2 \sum_\sigma \theta_\sigma^{(p,0)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma}\theta_\sigma^{(p,0)} d_\sigma^{(p-1)}/c_\sigma)^2 \\
&= 2\sum_\sigma \sum_k \alpha^{(p,k)} \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma \sum_k b_{i\sigma} \alpha^{(p,k)} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2 \qquad (21)\\
&= \sum_k \alpha^{(p,k)} (2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma) - \sum_i c_i (\sum_k \alpha^{(p,k)} \sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)})/c_\sigma)^2,
\end{aligned}$$

which by the Cauchy-Schwarz inequality

$$\begin{aligned}
&\geq \sum_k \alpha^{(p,k)} (2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma) \\
&\qquad - \sum_i c_i (\sum_k a^{(p,k)}) \{ \sum_k a^{(p,k)} (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2 \} \\
&= \sum_k \alpha^{(p,k)} \{ 2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2 \} \\
&\geq \sum_k \alpha^{(p,k)} (d_\tau^{(p-1)})^2/(c_\tau H) = (d_\tau^{(p-1)})^2/(c_\tau H).
\end{aligned}$$

A simpler and more convenient but somewhat more restrictive condition may be derived as a special case of (20). Let $\theta_\sigma^{(p)} = 0$ except for a set of one or

more σ so selected that $b_{i\sigma'}b_{i\sigma''} = 0$ for every i and every pair σ' and σ'' in the set. Then (20) becomes

$$\sum_{\sigma} \{2\theta_\sigma^{(p)} - (\theta_\sigma^{(p)})^2\} (d_\sigma^{(p-1)})^2/c_\sigma \geq (d_\tau^{(p-1)})^2/(c_\tau H). \tag{22}$$

Differentiating partially for a maximum with respect to one of the $\theta_\sigma^{(p)}$, we find that this special case of the condition will be satisfied if for one σ in the set, say π, such that

$$(d_\pi^{(p-1)})^2/c_\pi \geq (d_\tau^{(p-1)})^2/(c_\tau\sqrt{H}), \tag{23}$$

the value of $\theta_\pi^{(p)}$ is chosen in the range,

$$1/(2\sqrt{H}) \leq \theta_\pi^{(p)} \leq 2 - 1/(2\sqrt{H}) \tag{24}$$

and for every other σ in the set

$$0 \leq \theta_\sigma^{(p)} \leq 2, \tag{25}$$

all $\theta_\sigma^{(p)}$ not in the set being zero. A weighted average of such values of θ will satisfy (20) whence (6) and (7) follow.

In practice values of $\theta_\sigma^{(p)}$ satisfying (20) may be selected conveniently by the following procedure:

(a) Select a set of σ for at least one of which $\theta^{(p)}$ satisfies (23) and for every pair of which $b_{i\sigma'}b_{i\sigma''} = 0$. In so far as this restriction permits choose the σ corresponding to the larger values of $(d_\sigma^{(p-1)})^2/c_\sigma$.

(b) Determine values for each $\theta_\sigma^{(p)}$ in the set approximately equal to 1. Until other values are assigned to them assume all other $\theta_\sigma^{(p)} = 0$.

(c) Choose a σ not in the set, say ρ, for which $(d_\rho^{(p-1)})^2/c_\rho$ is relatively large and select a value for $\theta_\rho^{(p)}$ such that

$$\theta_\rho^{(p)} \doteq \{d_\rho^{(p-1)} - \sum_i \sum_{\sigma\neq\rho} c_i b_{i\rho} b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma\}/d_\rho^{(p-1)}. \tag{26}$$

(d) Having changed $\theta_\rho^{(p)}$ from 0 to a value approximately satisfying (26), continue with other σ not in the set letting ρ in (26) represent each in turn. The work may be terminated at any stage leaving some $\theta_\sigma^{(p)} = 0$.

7 Convergence of the adjustment. The condition equations may be written in the following form

$$\sum_i b_{i\sigma} d_i^{(0)} = d_\sigma^{(0)}, \tag{27}$$

as a system of consistent, but not necessarily independent, linear equations. They may also be written as conditions on the m_i. The least squares adjustment minimizes the quadratic form

$$S^{(0)} = \sum_i (d_i^{(0)})^2/c_i \tag{28}$$

subject to the restraints (27) Since the c_i are positive, $S^{(0)}$ is positive definite, and therefore a minimum exists and is non-negative. The values of the $d_i^{(0)}$ that minimize $S^{(0)}$ while satisfying (27) are $m_i - n_i$, the n_i being known and the m_i being the least squares adjusted values that are to be calculated.

If r is the rank of the matrix $\| b_{i\sigma} \|$, then from (15) and (16) it follows that r of the $d_i^{(0)}$ may be expressed as linear functions of the $t - r$ other $d_i^{(0)}$. The latter then constitute a set of $t - r$ independent variables. The normal equations

$$\partial S^{(0)} / \partial d_h^{(0)} = 0, \tag{29}$$

are obtained by differentiating $S^{(0)}$ with respect to each one of them in turn, one equation resulting for each value of h corresponding to a d_i in the set of independent variables. The normal equations (29) are a system of $t - r$ independent linear equations and can be written in the form

$$\Sigma \alpha_{i(h)} d_i^{(0)} = \Sigma \beta_{\sigma(h)} d_\sigma^{(0)}, \tag{30}$$

where the first summation is over the set of independent variables, and the second over the $d_\sigma^{(0)}$ in the r selected condition equations. The right-hand terms are constants. Since a least squares adjustment exists the equations are consistent and the rank of the matrix $\| \alpha_{i(h)} \|$ is $t - r$. Any $d_i^{(0)}$ in the set, say $d_{i'}^{(0)}$, is the quotient of two determinants the divisor being the determinant $| \alpha_{i(h)} |$ and the dividend being the determinant obtained by replacing the $\alpha_{i'(h)}$ by $\sum_\sigma \beta_{\sigma(h)} d_\sigma^{(0)}$. Consequently each $d_i^{(0)}$ whether in the set or not is a linear combination of the $d_\sigma^{(0)}$ and the sum of the absolute values of the coefficients of the $d_\sigma^{(0)}$ is finite. Therefore

$$\max | d_i^{(0)} / \sqrt{c_i} | \leq G \max | d_\sigma^{(0)} / \sqrt{c_\sigma} | \tag{31}$$

where G is $(\max c_\sigma / \min c_i)^{\frac{1}{2}}$ times the sum of the absolute values of the coefficients of the $d_\sigma^{(0)}$ in the linear combination for which such sum is a maximum. From (28)

$$S^{(0)} \leq t \max \{ (d_i^{(0)})^2 / c_i \} \leq G^2 t \max \{ (d_\sigma^{(0)})^2 / c_\sigma \} \tag{32}$$

whence

$$(d_\tau^{(0)})^2 / c_\tau \geq S^{(0)} / (G^2 t). \tag{33}$$

Consider now the discrepancies

$$d_i^{(p)} = m_i - m_i^{(p)} = d_i^{(p-1)} - c_i \sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)} / c_\sigma \tag{34}$$

between the m_i and the corresponding approximations $m_i^{(p)}$ and write the quadratic form

$$S^{(p)} = \sum_i (d_i^{(p)})^2 / c_i . \tag{35}$$

From (16), (18), and (34)

$$d_i^{(0)} = d_i^{(p)} + c_i \sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)}, \tag{36}$$

and

$$d_\sigma^{(0)} = d_\sigma^{(p)} + \sum_i \sum_\nu b_{i\sigma} b_{i\nu} c_i \lambda_\nu^{(p)}. \tag{37}$$

Hence the substitution of (36) in (27) merely changes (0) to (p) in the superscripts, the new equations being consistent by definition and the corresponding r of the $d_i^{(p)}$ being expressible as linear functions of the other $t - r$. Further (35) is positive definite and hence has a minimum, in fact substituting (36) in (28) we find that

$$\begin{aligned} \frac{\partial S^{(0)}}{\partial d_h^{(0)}} = \frac{\partial S^{(0)}}{\partial d_h^{(p)}} &= \frac{\partial}{\partial d_h^{(p)}} \{S^{(p)} + 2 \sum_i \sum_\sigma d_i^{(p)} b_{i\sigma} \lambda_\sigma^{(p)} + \sum_i c_i (\sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)})^2\} \\ &= \frac{\partial}{\partial d_h^{(p)}} (S^{(p)} + 2 \sum_\sigma d_\sigma^{(p)} \lambda_\sigma^{(p)}) = \frac{\partial S^{(p)}}{\partial d_h^{(p)}} = 0. \end{aligned} \tag{38}$$

Hence a least squares solution for the $d_i^{(p)}$ exists and it leads by (34) to the same values for the m_i as does the solution for the $d_i^{(0)}$. Since the coefficients $\alpha_{i(h)}$ and $\beta_{\sigma(h)}$ and the number G are functions of the $b_{i\sigma}$ and c_i they are invariant for the substitution. Consequently (30), (31), (32), and (33) may also be written with (p) in place of (0) in the superscripts (33) becoming

$$(d_r^{(p)})^2/c_r \geq S^{(p)}/(G^2 t). \tag{39}$$

From (20), (34), and (35)

$$\begin{aligned} S^{(p)} &= \sum_i (d_i^{(p)})^2/c_i \\ &= \sum_i (d_i^{(p-1)})^2/c_i - 2 \sum_i \sum_\sigma d_i^{(p-1)} b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma \\ &+ \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma)^2 \\ &= S^{(p-1)} - 2 \sum_\sigma \theta_\sigma^{(p)} (d_\sigma^{(p-1)})^2/c_\sigma + \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma)^2 \\ &\leq S^{(p-1)} - (d_r^{(p-1)})^2/(c_r H), \quad p > p' \end{aligned} \tag{40}$$

and from (39)

$$S^{(p)} \leq S^{(p-1)} - S^{(p-1)}/M \leq S^{(p')}\{1 - 1/M\}^{p-p'} \tag{41}$$

where

$$M = G^2 H/t. \tag{42}$$

Therefore, as $p \to \infty$, $p - p' \to \infty$, $S^{(p)} \to 0$, $d_i^{(p)} \to 0$, $m_i^{(p)} \to m_i$ and consequently the successive adjusted frequencies obtained by an iterative process in

which condition (20) is satisfied converge to the adjusted frequencies that are obtained by solving the normal equations

8. **Rate of convergence.** The computer is not as much interested in the proof of convergence as he is in how rapidly the successive adjustments reach a satisfactory degree of approximation. Equations (39) or (41) are of no help to him. The adjustment may be made in one step, with every $\theta = 1$, (a) if the condition equations are such that every $b_{i\sigma'} b_{i\sigma''} = 0$, $\sigma' \neq \sigma''$, i.e. if the adjustment can be separated into one-dimensional cases when redundant condition equations are ignored, or (b), in the two and three-dimensional cases, if the c_{ij} or c_{ijk} are proportional to the $c_{i.}$ and $c_{.j}$, or to the $c_{i..}$, $c_{.j.}$, $c_{..k}$ or $c_{ij.}$, $c_{i.k}$, and $c_{.jk}$ respectively. Except in these and other special cases, the rapidity of convergence depends on the $d_\sigma^{(0)}$ as well as on the $\| b_{i\sigma} c_i \|$ matrix. However, it seems that one can make very little use of the $d_\sigma^{(0)}$ to determine the rapidity of convergence without actually computing the successive adjustments or making some equivalent calculation.

Certain results can be obtained from the $\| b_{i\sigma} c_i \|$ matrix alone. Returning to the two-dimensional case and its notation, consider the matrix $\| c_{ij} \|$ and define

$$\delta_{ij} = c_{ij} - c_{i.} c_{.j}/c_{..}, \quad c_{..} = \sum_j c_{.j} \tag{43}$$

Let the adjustments be made with the restriction that $\theta_{i.}^{(p)} = 0$ and $\theta_{.j}^{(p)} = 1$ when p is even, and $\theta_{i.}^{(p)} = 1$ and $\theta_{.j}^{(p)} = 0$ when p is odd. Then if $p > 1$

$$\begin{aligned} d_{i.}^{(p)} &= -\sum_j (c_{ij}/c_{.j}) d_{.j}^{(p-1)} = \sum_j \sum_f (c_{ij}/c_{.j})(c_{fj}/c_{f.}) d_{f.}^{(p-2)} \\ &= \sum_j \sum_f (\delta_{ij}/c_{.j})(\delta_{fj}/c_{f.})\, d_{f.}^{(p-2)} \qquad (f = 1, 2, \cdots, r) \end{aligned} \tag{44}$$

The sum of the absolute values is

$$\sum_i | d_{i.}^{(p)} | \leq b_1^2 \sum_i | d_{i.}^{(p-2)} | \leq b_1^{p-2} \sum_i | d_{i.}^{(2)} | \tag{45}$$

where

$$b_1^2 = \sum_i \sum_j | \delta_{ij}/c_{.j} | \gamma_{.j} \tag{46}$$

$\gamma_{.j}$ being the greatest of the $| \delta_{ij}/c_{i.} |$ in the jth column. Similarly for $p > 2$

$$\sum_j | d_{.j}^{(p)} | \leq b_2^2 \sum_j | d_{.j}^{(p-2)} | \leq b_2^{p-2} \sum_j | d_{.j}^{(1)} | \tag{47}$$

where

$$b_2^2 = \sum_i \sum_j | \delta_{ij}/c_{i.} | \gamma_{i.} \tag{48}$$

$\gamma_{i.}$ being the greatest of the $| \delta_{ij}/c_{.j} |$ in the ith column.

Assume again the conditions just preceding (44). Let $u_{i.}$ be the minimum $c_{ij}/c_{.j}$ in the ith row. Likewise let $v_{.j}$ be the minimum $c_{ij}/c_{i.}$ in the jth column. Then since $\Sigma d_{i.}^{(p)} = \Sigma d_{.j}^{(p)} = 0$,

$$\Sigma \mid d_{.j}^{(p)} \mid = 2\Sigma^{+} d_{.j}^{(p)} = -2\Sigma^{-} d_{.j}^{(p)}, \tag{49}$$

the $+$ and $-$ signs indicating that the last two summations are over positive and negative values of $d_{.j}^{(p)}$ respectively. When p is even, of course, all values of $d_{.j}^{(p)} = 0$.

From (44)

$$\begin{aligned} d_{i.}^{(p)} &= -\sum_j c_{ij} d_{.j}^{(p-1)}/c_{.j} = \sum_j^{-} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - \sum_j^{+} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} \\ &= \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - 2\sum^{+} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} \\ &= 2\sum_j^{-} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} \end{aligned} \tag{50}$$

and by (49)

$$\mid d_{i.}^{(p)} \mid \leq \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - u_{i.} \sum_j \mid d_{.j}^{(p-1)} \mid , \tag{51}$$

$$\sum_i \mid d_{i.}^{(p)} \mid \leq \sum_j \mid d_{.j}^{(p-1)} \mid (1 - \sum_i u_{i.}). \tag{52}$$

Similarly

$$\sum_j \mid d_{.j}^{(p)} \mid \leq \sum_i \mid d_{i.}^{(p-1)} \mid (1 - \sum_j v_{.j}) \tag{53}$$

Let $b_3 = 1 - \sum_i u_{i.}$ and $b_4 = 1 - \sum_j v_{.j}$, then

$$\sum_i \mid d_{i.}^{(p)} \mid \leq b_3 b_4 \sum_i \mid d_{i.}^{(p-2)} \mid \leq (b_3 b_4)^{\frac{1}{2}p-1} \sum_i \mid d_{i.}^{(2)} \mid . \tag{54}$$

Now b_3 or b_4 may be greater or less than b_1 or b_2 but, unlike b_1 and b_2, they can not exceed unity. Let b^2 be the lesser of b_1^2 and $b_3 b_4$. Then under the conditions stated with equation (44)

$$\Sigma \mid d_{.j}^{(p+1)} \mid \leq \Sigma \mid d_{i.}^{(p)} \mid \leq b^2 \Sigma \mid d_{i.}^{(p-2)} \mid \leq b^{p-2} \Sigma \mid d_{i.}^{(2)} \mid \leq b^{p-2} \Sigma \mid d_{.j}^{(1)} \mid . \tag{55}$$

It follows from (40) that

$$\begin{aligned} S^{(p)} &= S^{(p+1)} + \sum_i (d_{i.}^{(p)})^2/c_{i.} + \sum_j (d_{.j}^{(p)})^2/c_{.j} \\ &= \sum_{h=p}^{\infty} \{\sum_i (d_{i.}^{(h)})^2/c_{i.} + \sum_j (d_{.j}^{(h)})^2/c_{.j}\} \\ &\leq \sum_{h=p}^{\infty} \{(\sum_i \mid d_{i.}^{(h)} \mid)^2/\min c_{i.} + (\sum_j \mid d_{.j}^{(h)} \mid)^2/\min c_{.j}\} \\ &\leq (\sum_i \mid d_{i.}^{(p)} \mid + \mid d_{i.}^{(p-1)} \mid)^2 \{(1/\min c_{i.}) + (1/\min c_{.j})\}/(1 - b^4). \end{aligned} \tag{56}$$

The reduction in $S^{(p)}$ in g steps of the iterative process is

$$D = S^{(p)} - S^{(p+g)} = \sum_{h=p}^{p+g-1} [\sum_i (d_{i.}^{(h)})^2/c_{i.} + \sum_j (d_{.j}^{(h)})^2/c_{.j}] \tag{57}$$
$$\geq \sum_{h=p}^{p+g-1} [(\sum_i | d_{i.}^{(h)} |)^2/(r \max c_{i.}) + (\sum_j | d_{.j}^{(h)} |)^2/(s \max c_{.j})].$$

from which, by (55), if $g > 1$ is odd,

$$D \geq \frac{1 - b^{-4g}}{1 - b^{-4}} (\sum_i | d_{i.}^{(p+g)} | + | d_{i.}^{(p+g+1)} |)^2 \left(\frac{1}{r \max c_{i.}} + \frac{1}{s \max c_{.j}}\right). \tag{58}$$

The relative decrease in $S^{(p)}$ is, therefore, by (56),

$$\frac{D}{S^{(p)}} = \frac{D}{D + S^{(p+g)}} \geq \left\{1 + \frac{1/\min c_{i.} + 1/\min c_{.j}}{b^4 (b^{-4g} - 1) \left(\frac{1}{r \max c_{i.}} + \frac{1}{s \max c_{.j}}\right)}\right\}^{-1}. \tag{59}$$

If the g steps actually have been taken a better lower limit for the relative decrease in $S^{(p)}$ may be obtained by computing D from (57) and using (56) for $S^{(p+g)}$. Similar equations can be written using b_2.

These results can be shown to be valid for an adjustment in which $\theta_{i.}^{(p)} = \theta_{.j}^{(p)} = 1$ at the first and any of the subsequent steps. They also can be extended to the three-dimensional cases but not to three-dimensional adjustments with every $\theta = 1$.

9. **Improvement resulting from the adjustment.** The least squares adjustment eliminates a portion of the errors of sampling, i.e a portion of χ^2, from the set of frequencies estimated from the sample. In fact any adjustment that satisfies the condition equations does this.

Let ϵ_i be the error in the ith value given by the sample and $\delta_i^{(p)}$ the error in the pth approximation to the least squares adjusted value. Then

$$\delta_i^{(p)} = \epsilon_i + c_i \sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)}, \tag{60}$$

and

$$\sum_i (\delta_i^{(p)})^2/c_i = \sum_i \epsilon_i^2/c_i + 2 \sum_\sigma \lambda_\sigma^{(p)} \delta_\sigma^{(p)} - \sum_i c_i (\sum_\sigma b_\sigma \lambda_\sigma^{(p)})^2. \tag{61}$$

The complete adjustment makes $\delta_\sigma^{(p)}$ vanish and therefore, since the last term is non-negative, $\Sigma\delta_i^2/c_i < \Sigma\epsilon_i^2/c_i$ except in the trivial case in which all $d_\sigma^{(0)} = 0$. From (37)

$$\sum_i (\delta_i^{(p)})^2/c_i = \sum_i \epsilon_\sigma^2/c_i + \sum_\sigma \lambda_\sigma (d_\sigma^{(p)} - d_\sigma^{(0)}). \tag{62}$$

The last term may be computed readily at any stage in the iteration. If the sampling is at random, $k\, \Sigma\epsilon_i^2/c_i$ is distributed approximately as χ^2 with $t - 1$ degrees of freedom, where k is the ratio of the c_i to the corresponding error

variances of the n_{ij}. Therefore it would seem appropriate to compute $k\ \Sigma\lambda_\sigma d_\sigma^{(0)}$, the reduction in χ^2, as a measure of the improvement achieved in the final adjustment.

10. The method of iterative proportions. The iterative proportions method described in the earlier paper [1] implicitly defines, in the two dimensional case,

$$m_{ij} = \mu_{i.}\,\mu_{.j}\,n_{ij}, \tag{63}$$

the $\mu_{i.}$ and $\mu_{.j}$ being given by the $r + s$ condition equations

$$m_{i.} = \sum_j \mu_{i.}\,\mu_{.j}\,n_{ij}, \qquad m_{.j} = \sum_i \mu_{i.}\,\mu_{.j}\,n_{ij}, \tag{64}$$

any $r + s - 1$ of which constitute a consistent system of independent equations in $r + s$ unknowns. One multiplier, say $\mu_{1.}$, may be fixed arbitrarily. Then for a $2 \times s$ table it is necessary to solve an equation of the sth degree. If $s = 2$, there is only one acceptable solution, given by the positive root; if $s = 3$, there is only one solution of the cubic for which all the adjusted frequencies are non-negative. For 3×3 and larger tables the adjustment appears to involve the solution of equations of the tenth or higher degree and there is then no choice but to use methods of approximation.

The adjusted frequencies given by the method of iterative proportions are not identical to those given by the method of least squares. When the adjustments are small relative to the frequencies adjusted, however, the results given by this method approximate those of least squares. For the two-dimensional case the successive adjustments converge to a set of frequencies that satisfy the condition equations The author has not found a proof of convergence or divergence for more than two dimensions.

I wish to express my appreciation of many stimulating conversations with Dr. W. Edwards Deming on this and related problems, and of the helpful critical reading of certain portions of the manuscript by Dr. Joseph F. Daly.

REFERENCE

[1] W Edwards Deming and Frederick F Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *Annals of Math Stat*, Vol. 11 (1940), p 427.

ESTIMATION OF VOLUME IN TIMBER STANDS BY STRIP SAMPLING

BY A. A. HASEL

California Forest and Range Experiment Station[1]

1. Introduction. The present paper is the second of a proposed series, in which it is intended to present a systematic study of the properties of several methods of sampling timber stands and statistical treatments of the samples

The effects of size, shape, and arrangement of sampling units on the accuracy of sample estimates of timber stand volume were reported in the earlier paper [1] for 5,760 acres of the Blacks Mountain Experimental Forest. With complete inventory data, the nature of stand variation was shown to be such that 2.5-acre plots, the smallest size tested, were more efficient sampling units than larger plots, i.e., for a given intensity of sampling the sampling error was smaller. Long, narrow plots were more efficient than square plots of the same size. Line-plot sampling units consisting of two or more equally spaced plots along lines of fixed length were as efficient as single-plot sampling units and more efficient than strips consisting of plots contiguous end to end. Improvement in the accuracy of estimates was obtained by subdividing the area into rectangular blocks of equal size, and sampling each block to the same intensity. By systematic sampling, whereby the center lines of parallel line-plot or strip sampling units were spaced equidistant, the sample estimates of stand volume were improved over estimates from comparable random samples. Treatment of the volumes on individual plots of systematic samples as random sampling observations, however, as is sometimes done in practice, was shown to give seriously biased estimates of sampling error

In the present paper we shall be concerned with sample estimates from strip samples taken within blocks of irregular shape, and consequently with sampling units which vary in length within samples The methods will be equally applicable to line plot samples.

Following the general ideas expressed by Neyman [2] it is felt that. (1) If the formulae of the theory of probability have to be applied at all to the treatment of samples, the theoretical model of sampling must involve some element of randomness (2) This element of randomness may conveniently be introduced by a random selection of the sample, but may also be assumed present in the distribution of deviations of timber stand volumes in the area sampled from a postulated pattern. (3) Many attempts to treat systematic arrangements statistically are faulty because the treatment consists in applying to systematic arrangements formulae that are deduced under the assumption of randomness. If the arrangement of sampling is a systematic one, and random errors are

[1] Maintained by the U. S Department of Agriculture at Berkeley, in cooperation with the University of California.

ascribed to Nature, then the treatment of the data should be based on formulae deduced under explicit assumption of the systematic arrangement of sampling and of some random element in the material. An example of this kind of treatment is provided by Neyman's method of parabolic curves [2] devised for the treatment of systematically arranged agricultural experiments. (4) Lastly, a mathematical treatment of any practical problem is useful only if the predictions of the theory are in satisfactory agreement with the empirical facts. Whether the method of sampling is random or systematic, the mathematical theory of sampling always involves certain elements that are postulated, either in respect to the method of sampling itself or in respect to the material sampled. To have a reasonable certainty that a particular mathematical treatment is useful in practice it is necessary to make empirical studies to find out whether the deviations from postulates of the theory that may occur in actual situations do or do not seriously affect the validity of the predictions.

2. Notation and definitions. Before proceeding to the main subject of this paper it may be useful to explain the meaning of certain statistical terms and symbols. Following Neyman, a sharp distinction is made between three different conceptions that are frequently confused by the practical statistician.

Definition 1: If $u_1, u_2, \cdots, u_N$ are any fixed numbers, whether provided by some already completed experiment involving randomness, or just arbitrarily selected, Karl Pearson's term "standard deviation" of these numbers and the letter S will be used to denote the expression $S = \sqrt{\Sigma(u_i - \bar{u})^2/N}$ in which $\bar{u} = \Sigma u_i/N$ is the mean of the u's.

Now let X denote a random variable, that is a variable the value of which is going to be determined by a chance experiment. Thus X may be the timber volume on a strip that is going to be selected at random from an area. Denote by $E(X)$ the mathematical expectation of variable X capable of possessing values $u_1, u_2, \cdots, u_n$. Then

$$E(X) = u_1p_1 + u_2p_2 + \cdots + u_np_n,$$

in which the p's are the respective probabilities of all possible different values of X.

Definition 2: The words "standard error of X" and the letter σ_x will be used to denote the expression

$$\sigma_x = \sqrt{E[X - E(X)]^2}.$$

It will be noticed that the standard error of a random variable X may have its value equal to the standard deviation of some numbers u but that this does not mean that the two conceptions are identical or even similar. The $E(X)$, and consequently σ_x, can be calculated only when the probability law of X is known, and are constant for the population from which samples are drawn. On the other hand, S can be calculated for any sample of the population and changes in value from one sample of u's to another.

Before proceeding to the third conception, that of an estimate of the standard error, which is occasionally confused with the standard deviation or the standard error, the unbiased estimate of a parameter must be defined [3].

Consider a set of n random variables $X_1, X_2, \cdots, X_n$. These may be, for example, the volumes of timber to be observed on n strips that are going to be selected from some area by one random method or another. Denote by θ a parameter involved in the probability law of the X's. For example, θ may be the total volume of timber in the area.

Let F be any function of the X's.

DEFINITION 3: If it happens that the mathematical expectation of F is identically equal to θ, then it will be said that F is an unbiased estimate of θ.

Usually there will be an infinity of unbiased estimates of a parameter θ. They may be classified by the nature of the function F. Thus linear estimates may be considered such that

$$F = \lambda_0 + \lambda_1 X_1 + \cdots + \lambda_n X_n$$

in which the λ's stand for some fixed numbers.

DEFINITION 4: It will be said that a linear unbiased estimate of θ is the best linear unbiased estimate (B. L. U. E.) if its standard error is smaller than or, at most, equal to that of any other linear unbiased estimate.

It happens frequently that, while it is possible to determine the best linear unbiased estimate F of a parameter, it is not possible to calculate the value of its standard error, σ_F. For this purpose it would be necessary to know the whole population sampled. In such cases an unbiased estimate of the square, σ_F^2, is calculated. An unbiased estimate of the square of the standard error of F will be denoted by μ_F^2. This is the third of the conceptions mentioned above.

The reason for the extensive use of the linear unbiased estimates and of their standard errors considered as measures of accuracy is the so-called Theorem of Liapounoff. Its content can roughly be explained as follows: If the variables $X_1, X_2, \cdots, X_n$ are independent and the number n not too small, then the probability that $F - \theta$ will exceed a fixed multiple of σ_F is approximately equal to the probability as determined by the normal law. The above conclusion remains true whatever the probability distribution of the X's that is likely to be met in practice and also in certain cases where the X's are mutually dependent, for example, when they are determined by sampling a finite population without replacement [4].

The above conclusions do not apply to estimates that are biased in the sense of the above definition. . Also the standard error of such an estimate would not be a satisfactory measure of its accuracy.

3. Description of data. Complete inventory data from the Blacks Mountain Experimental Forest, located in the Lassen National Forest, provide suitable material for testing the applicability of sampling theory to timber cruising.

The timber is a virgin, all-aged stand, classed as pure pine type, with more than 90 per cent of the volume in ponderosa pine and Jeffrey pine. Most of the volume is in over-mature trees, i.e., trees over 300 years in age. The stand is considered to be fairly representative of the medium and the poor site qualities of the northeastern California plateau.

With the exception of a few localities, all of the area was mapped as of uniform timber type according to the standards commonly used. Being fairly uniform also with respect to site quality, it may therefore be considered as a single *stratum*. Variability of stand volume from place to place within a stratum may be generally expected to be less, on the average than variability between places in different strata. Likewise, within a stratum, variability within compact subdivisions may be expected to be less than average variability within the whole. Heterogeneity can therefore be controlled somewhat by subdividing the stratum into blocks and treating each block as a separate population.

More frequently than not, in practice, volume estimates are needed both for the total timbered area and for separate working units or compartments within the area. In general, working unit boundaries are defined by roads, ridge tops, drainage channels, and regular land subdivision lines. These working units can be taken conveniently as blocks, or if large enough, may be subdivided into two or more blocks. Such is the basis used for subdividing the area in the present study.

The complete inventory data for these blocks are given in Table I. All the strips are $2\frac{1}{2}$ chains in width and extend in an east-west direction. The length, X, is given in 10-chain units, and the volume, Y, is given in units of 1,000 feet board measure.

4. Method of estimation based on correlation between volume and strip length. The usual practice in sampling timber stand volume is to take measurements on plots or strips that are either regularly spaced or selected at random from all possible plots or strips within blocks. Oftener than not blocks are irregular in shape, and the number of plots along lines or the lengths of strips will vary. This variation introduces the matter of proper "weighting" in calculating sample statistics. Such is the case in 15 of the 20 Blacks Mountain blocks.

If we let Y_i represent the volume on the ith strip of length X_i, with length expressed say in 10-chain units, and assume that the entire block contains a population of N strips, then the average volume to the unit of strip is $\beta = \sum_{i=1}^{N} Y_i \Big/ \sum_{i=1}^{N} X_i$. It is obvious that, if $\sum X_i$ is known, and this is assumed to be true, the problem of estimating β is equivalent with that of estimating the total volume. The usual procedure of estimating is this:

Out of the N strips within the block a sample of n is taken, giving n pairs of numbers selected out of the X's and Y's. Let us denote them by

$$x_1, y_1; x_2, y_2; \cdots; x_n, y_n.$$

The ratio $b = \sum_{i=1}^{n} y_i \Big/ \sum_{i=1}^{n} x_i$, is then considered an estimate of β, so that the estimate of the total volume in the block is $b \sum_{i=1}^{N} X_i$.

Our purpose now will be to study the above estimate b from the point of view of unbiasedness In this paper it is assumed that the sampling of strips is purely random. To find out whether b is an unbiased estimate or not, its expectation must be calculated. This will be done in two steps. To begin with assume that the *values* $x_1, x_2, \cdots, x_n$ are chosen in one way or another and fixed. The value of b will then depend on the y's only. It is possible that to a given value of x, say x_1, there will correspond just one value of y_1 in the block, but generally there will be several strips of the same length x_1 with varying volumes of timber The selection of any strip of this group to be included in the sample will keep the denominator of b constant, but will cause some variation in the numerator The expectation of b calculated under the assumption that the x's are fixed is

$$E(b \mid x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} E(y_i \mid x_i) \Big/ \sum_{i=1}^{n} x_i, \tag{1}$$

in which $E(y_i \mid x_i)$ denotes the expectation of y_i calculated under the assumption that x_i has a fixed value Obviously $E(y_i \mid x_i)$ will be what is called the regression function of y on x, or of volume on the length of strip.

It is safe to say that the graph of $E(y \mid x)$ would almost always be rather irregular. On the other hand, it is known that the substitution of smooth curves representing the regressions for the true irregular polygons frequently gives results that are surprisingly accurate Therefore it would not be unreasonable to use the assumption that $E(y \mid x)$ can be represented by a polynomial of some moderate degree,

$$E(y \mid x) = A_0 + A_1 x + A_2 x^2 + \cdots + A_s x^s.$$

Substitution of this expression in (1) gives

$$E(b \mid x_1, x_2, \cdots, x_n) = \frac{nA_0}{\sum_{i=1}^{n} x_i} + A_1 + A_2 \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i} + \cdots + A_s \frac{\sum_{i=1}^{n} x_i^s}{\sum_{i=1}^{n} x_i}.$$

But this is the conditional expectation of b, calculated under the assumption of fixed x's, is only an intermediate stage in the calculations. We need an absolute expectation, calculated under the assumption that the x's are selected at random. This gives

$$E(b) = A_0 E\left(\frac{n}{\sum_{i=1}^{n} x_i}\right) + A_1 + A_2 E\left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i}\right) + \cdots + A_s E\left(\frac{\sum_{i=1}^{n} x_i^s}{\sum_{i=1}^{n} x_i}\right) \tag{2}$$

TABLE I

Complete inventory data for 15 blocks of the Blacks Mountain Experimental Forest

Strip[1] number	Block number 1		2		3		4		5		6		8		9	
	x	y	x	y	x	y	x	y	x	y	x	y	x	y	x	y
1	12	762.4	9	470.6	4	63.6	7	174.3	5	121.4	9	331.1	7	247.5	3	143.6
2	12	651.0	9	426.1	4	73.7	7	159.2	5	169.5	9	377.7	7	286.4	3	159.2
3	12	461.4	9	448.5	4	91.7	7	156.6	6	315.8	9	295.1	7	339.5	4	209.2
4	12	521.1	9	401.5	4	61.4	7	139.6	7	307.1	9	237.8	7	360.2	4	209.5
5	12	652.7	9	372.1	4	35.5	7	198.0	7	318.9	9	305.0	8	336.2	4	227.2
6	12	543.7	9	372.2	4	82.8	7	168.7	7	366.0	9	284.4	8	332.7	5	247.0
7	12	541.5	9	410.6	4	109.2	7	127.1	9	445.0	9	362.9	8	330.1	5	277.0
8	12	589.6	9	322.8	4	109.6	7	181.3	9	406.4	9	352.4	8	403.6	5	371.9
9	11	532.6	9	380.6	5	114.9	7	155.8	9	427.1	9	345.1	8	378.9	6	303.5
10	11	516.9	9	429.9	5	101.6	7	207.3	9	448.7	9	354.0	6	279.5	6	207.7
11	11	519.5	9	434.1	5	94.4	7	181.4	9	381.6	9	401.2	4	206.5	7	305.0
12	11	538.8	9	394.5	5	104.3	7	121.3	9	243.3	9	381.0	2	130.1	7	353.4
13	10	508.6	9	542.9	6	161.2	7	124.6	9	360.0	8	408.4	1	45.3	8	368.7
14	10	448.6	9	606.6	6	191.6	7	86.0	9	343.9	8	374.5			8	361.1
15	10	492.5	8	416.5	6	223.5	3	115.3			8	374.3			8	332.8
16	10	498.3	8	325.9	6	164.8	3	128.8			8	337.3			8	289.6
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																
32																
Total	180	8,779.2	142	6,755.4	76	1,783.8	104	2,425.3	109	4,654.7	140	5,522.2	81	3,676.5	91	4,366.4

TABLE I (*Continued*)

Strip number	Block number													
	12		13		14		15		18		19		20	
	x	y	x	y	x	y	x	y	x	y	x	y	x	y
1	1	26.1	1	21.1	6	184.1	10	300.7	8	286.0	2	34.4	2	104.8
2	1	48.5	1	23.3	6	198.1	9	170.1	8	306.1	2	43.7	2	104.3
3	4	137.8	2	40.4	7	240.9	9	270.7	8	319.0	2	33.3	1	20.7
4	8	329.9	3	43.1	7	348.4	9	386.7	8	285.5	2	57.0	1	36.7
5	10	354.8	4	50.5	7	315.3	7	246.7	8	282.4	3	89.1	1	56.1
6	12	401.0	4	95.9	7	262.4	7	215.8	8	281.8	3	72.1	1	42.4
7	14	499.7	5	145.4	7	243.6	7	248.4	7	218.8	3	43.0	1	22.9
8	15	554.5	5	155.6	7	312.7	7	265.9	7	244.2	3	81.4	1	75.5
9	15	534.0	5	126.9	7	287.2	7	292.4	6	171.7	3	91.0	1	62.5
10	15	519.0	5	161.0	7	254.4	7	221.6	5	127.5	3	94.3	1	39.5
11	15	570.7	5	213.9	7	303.8	7	268.4	2	58.4	3	141.4	1	40.8
12	9	402.8	5	143.0	7	338.9	7	291.3	2	79.9	3	107.6	1	72.8
13	9	354.5	5	201.0	7	303.3	7	279.0	1	28.0	3	108.8	1	34.5
14	9	363.1	5	198.3	7	233.6	7	312.6	1	13.8	3	117.3	1	35.7
15	9	561.9	5	173.1	7	284.6	7	303.2			3	122.3	1	22.3
16	6	223.8	5	219.9	7	306.4	7	278.5			3	97.5	1	52.1
17	6	245.5	5	296.8	7	298.4	7	301.1			3	126.8	1	37.0
18	5	255.7	5	150.5	7	218.7	7	260.3			3	163.0	1	36.5
19	5	244.8	5	157.5	6	230.6	7	258.5			3	111.9	2	61.0
20	5	245.1	5	162.6	6	187.8	7	261.6			3	140.8	2	72.2
21	4	248.0	5	140.7	6	189.4	7	256.0			2	87.9	2	95.1
22	4	250.3	5	212.1	6	179.8	6	178.2			2	89.0	2	99.3
23	4	200.4	5	142.8	6	124.6	6	242.9					2	79.8
24	2	98.1	5	224.5	5	151.5	4	193.1					2	82.1
25	2	151.8	6	174.0	3	109.2	3	149.8					2	56.7
26	1	40.8	6	187.2	2	72.8	2	58.5					2	77.3
27	1	45.8			1	31.5	1	36.9					2	68.5
28													2	73.4
29													2	31.5
30													2	33.6
31													2	44.5
32													2	105.5
Total	191	7,908.4	117	3,861.1	165	6,212.0	178	6,548.9	79	2,703.1	60	2,053.6	48	1,877.6

[2] Numbered in order from north to south within blocks

The value of β has the form of (2), except that instead of $E(\Sigma x_i^n/\Sigma x_i)$ it contains $\Sigma X_i^n/\Sigma X_i$. Since in general the former does not necessarily equal the latter, for the unbiasedness of b it is necessary and sufficient that $A_0 = A_2 = \cdots = A_s = 0$ This condition implies that the regression line of y on x is a straight line and passes through the origin of coordinates,

$$E(y \mid x) = A_1 x. \tag{3}$$

Whether (3) is satisfied is a question of fact and can be authoritatively answered only by direct studies of regressions on some extensive inventory data. It may be noted also that in order to presume that (3) is *usually* satisfied, it should be established for a large number of areas. On the contrary, if a study of only a few areas shows that (3) is not true, then it would not be wise to take it for granted when attempting to make a sampling inventory of an unfamiliar area.

To investigate this point, linear regression equations of volume on the length of strip were calculated for 15 blocks of the Blacks Mountain Experimental Forest and it was found that the constant terms were both positive and negative with their absolute values varying from 12 to 677. The conclusion drawn is that the usual estimate b of the average volume per unit of strip is likely to be biased and that there is justification in looking for an alternative method leading to unbiased estimates.

5. Best linear unbiased estimate of volume, based on the linear regression of volume on length of strip. In this section will be suggested a method of estimating the total volume, say θ, of a timber stand, which could be considered as an improvement on the one considered above. The new method consists of using a linear unbiased estimate of θ. In order to deduce the form of this estimate, certain assumptions have to be taken for granted concerning the timber stands to be sampled, and if it happens that these assumptions are unsatisfied in a particular case, the new estimate will not necessarily possess the desired property of unbiasedness

In deducing the estimate F it will be assumed that the timber stand to be sampled satisfies the following conditions: (1) That the regression of timber volume on length of strip, X, be (approximately) linear and (2) that the variability of the Y's for a fixed X is precisely known. It will not be assumed, however, that the linear regression line passes through the origin of coordinates, and this will allow F to be unbiased in such cases, as exhibited above, where b is biased. Following the Markoff method [5], [3] it can easily be shown that there is an infinity of linear estimates of θ which are unbiased under condition (1) It follows that a choice can be made among them so as to diminish the standard error. This, however, is possible only when something is known about the variability of the Y's when the value of X is fixed. For the present we shall assume condition (2) concerning this particular point, but in practice this will generally be quite impossible. This point will be considered further in Section 6.

Consider then a sure or non-random variable[3] X able to assume the particular values $X_1, X_2, \cdots, X_s$. Assume that there is a finite population π_i of N_i numbers $u_{i1}, u_{i2}, \cdots, u_{iN_i}$ corresponding to each value X_i, $i = 1, 2, \cdots, s$. Assume that the mean u_i of the population π_i is a linear function of X_i, i.e., for any i,

$$u_i = A + BX_i,$$

with some unknown values of A and B.

Assume that out of each population π_i there is selected without replacement a random sample of n_i individuals, with $0 \leq n_i \leq N_i$, and denote by $y_{i1}, y_{i2}, \cdots, y_{in_i}$ the values of the u's to be drawn

If the regression of the amount of timber on the length of the strip is linear, then the problem of estimating the total stand is equivalent to that of estimating

$$\theta = \sum_{i=1}^{s} N_i(A + BX_i) = A \sum_{i=1}^{s} N_i + B \sum_{i=1}^{s} N_i X_i .$$

Since the length of the strip, X, could be measured from any arbitrarily chosen origin, no generality will be lost by assuming $\sum_{i=1}^{s} N_i X_i = 0$, so that $\theta = A \sum_{i=1}^{s} N_i = AN$ (say). Weighting the y_{ij} equally for each fixed i the B. L. U. E. of θ may be denoted by

$$F = \sum_{i=1}^{s} n_i \lambda_i y_{i.} , \tag{4}$$

in which $y_{i.} = {}_j\Sigma y_{ij}/n_i$. Here the λ's must satisfy the conditions of unbiasedness,

$$E(F) \equiv \theta, \tag{5}$$

and of optimum,

$$\sigma_F^2 = \text{minimum}. \tag{6}$$

It may easily be shown that condition (5) will be fulfilled by (4) if the λ_i's are so selected that

$$\sum_{i=1}^{s} n_i \lambda_i = N; \qquad \sum_{i=1}^{s} n_i \lambda_i X_i = 0. \tag{7}$$

[3] This is an English translation from an excellent French term "nombre certain" and "fonction certaine" to denote a non-random number and non-random function, invented by Fréchet.

Condition (6) may now be considered. From the general formula for the variance of a linear function of several random variables and the fact that y_{ij} is independent of y_{kl},

$$
\begin{aligned}
\sigma_\lambda^2 &= \sum_{i=1}^{s} \{n_i\lambda_i^2 S_i^2 + n_i(n_i - 1)\lambda_i^2 E[(y_{ik} - u_{i\cdot})(y_{il} - u_{i\cdot})]\} \\
(8) \qquad &= \sum_{i=1}^{s} \left[n_i\lambda_i^2 S_i^2 - \frac{S_i^2}{N_i - 1}(n_i^2\lambda_i^2 - n_i\lambda_i^2)\right] \\
&= \sum_{i=1}^{s} S_i^2 \frac{n_i(N_i - n_i)}{N_i - 1}\lambda_i^2 = \sum_{i=1}^{s} A_i^2\lambda_i^2 \text{ (say)},
\end{aligned}
$$

in which S_i^2 stands for the $(\text{S.D.})^2$ of the population π_i, i.e.,

$$S_i^2 = \frac{1}{N_i}\sum_{j=1}^{N_i}(u_{ij} - u_{i\cdot})^2.$$

In addition to satisfying equations (7), the λ_i's must be selected so as to minimize (8).

Using the method of Lagrange, we find

$$\lambda_i = \frac{n_i}{A_i^2}(\alpha + \beta X_i), \tag{9}$$

for the case where $0 < n_i < N_i$ and $A_i \neq 0$. If $n_i = N_i$, then $A_i = 0$ and $\alpha + \beta X_i = 0$.

Assume first that all $n_i < N_i$, $i = 1, 2, \cdots, s$. Then α and β are obtained from equations (7) after substituting in them (9), namely

$$
\begin{cases}
\alpha \sum_i w_i + \beta \sum_i w_i X_i = N \\
\alpha \sum_i w_i X_i + \beta \sum_i w_i X_i^2 = 0,
\end{cases} \tag{10}
$$

where, for simplicity

$$w_i = \frac{n_i^2}{A_i^2} = \frac{(N_i - 1)n_i}{(N_i - n_i)S_i^2}; \qquad \sum_{i=1}^{s} w_i = W. \tag{11}$$

If w_i is considered as the weight of the observations at $X = X_i$, it will be convenient to introduce a weighted mean and weighted S.D. of X's as follows:

$$\bar{x} = \frac{\sum_{i=1}^{s} w_i X_i}{W}; \qquad S_x^2 = \frac{\sum_{i=1}^{s} w_i X_i^2}{W} - \bar{x}^2. \tag{12}$$

With this notation equations (10) can be rewritten and easily give

$$
\begin{cases}
\beta = -\dfrac{N\bar{x}}{WS_x^2}, \\
\alpha = \dfrac{N(S_x^2 + \bar{x}^2)}{WS_x^2}.
\end{cases} \tag{13}
$$

Substituting these values into (4), simple transformations give

$$F = N(\bar{y} - \bar{x}b_0), \tag{14}$$

in which $\bar{y} = \sum_i w_i y_i / W$, and b_0 represents the unbiased estimate of B and is given by

$$b_0 = [(1/W) \sum_i w_i X_i y_i - \bar{x}\bar{y}]/S_x^2.$$

The next step is to calculate σ_F^2. Substituting (9) in (8) and using (11) and (12) gives

$$\sigma_F^2 = W(\alpha + \beta\bar{x})^2 + W\beta^2 S_x^2.$$

Using (13) gives finally

$$\sigma_F^2 = \frac{N^2}{W}\left(1 + \frac{\bar{x}^2}{S_x^2}\right). \tag{15}$$

If X is the length of a given strip in any chosen units and $\bar{X}$ the average of such X's for a given block, then (14) and (15) may be written

$$\begin{cases} F = N[\bar{y} + b_0(\bar{X} - \bar{x})] \\ \sigma_F^2 = \dfrac{N^2}{W}\left[1 + \dfrac{(\bar{X} - \bar{x})^2}{S_x^2}\right]. \end{cases} \tag{16}$$

Similarly for the case where one of the n_i's equals N_i, for example, $n_1 = N_1$, we find

$$F = N[y_{1.} + \bar{b}(\bar{X} - X_1)], \tag{17}$$

in which

$$\bar{b} = \frac{\sum_{i=2}^{s} w_i(X_i - X_1)(y_i - y_{1.})}{\sum_{i=2}^{s} w_i(X_i - X_1)^2}.$$

Also

$$\sigma_F^2 = \frac{N^2(\bar{X} - X_1)^2}{\sum_{i=2}^{s} w_i(X_i - X_1)^2}. \tag{18}$$

It should be emphasized that X_1 in the above formulae does not necessarily represent the smallest of the X's but the one of them for which $n_i = N_i$.

The case where two or more of the n_i's are respectively equal to the corresponding N_i's need not be considered in detail. Together with the assumption of a strict linearity of regression such an assumption, for example, that $n_1 = N_1$, and $n_2 = N_2$, would lead to the conclusion that the regression of volume on the length of strip is accurately known and that the estimation of θ could be made

without error. Owing to the fact that the hypothesis about the linearity of regression is, at best, only approximately correct, the errors of estimation will always be present and it is imperative either to arrange the sampling so as to have at most one of the n_i's equal to the corresponding N_i, or to base the statistical treatment of the sample on a theory different from the one considered here.

6. Additional hypotheses concerning S_i^2. The formulae (16) with the n_i's determined by (11) are impossible to apply in practice because we do not know the values of the S_i^2. The best we can do is to make plausible guesses as to what may be the values of the S_i^2. These guesses are bound to be at most approximately correct and therefore the estimates of θ that one can apply in practice will be only "approximately best." It is easy to see, however, that we may keep them unbiased.

Suppose that we denote by r_i^2 the presumed value of S_i^2. Substituting this value in place of S_i^2 in (8) we should repeat all the calculations, leading us to such λ_i' that will assure the unbiasedness of, say

$$F_0 = \sum_{i=1}^{s} n_i \lambda_i' y_{i\cdot},$$

and also a minimum value of, say

$$\sigma_0^2 = \sum_{i=1}^{s} r_i^2 \frac{n_i(N_i - n_i)}{N_i - 1} \lambda_i'^2.$$

The values of the λ_i' will be obtained from the same formulae as those of λ_i, except that instead of S_i^2 they will depend on r_i^2. Consequently F_0 will have the same form as F,

$$F_0 = N[\bar{y}' + b_0'(\bar{X} - \bar{x}')], \tag{19}$$

with the difference that $\bar{x}'$, $\bar{y}'$, S_x', and b_0' will now have to be calculated with different weights, say

$$v_i = \frac{(N_i - 1)n_i}{(N_i - n_i)r_i^2}; \qquad V = \sum_{i=1}^{s} v_i.$$

If the form of the unbiased estimate F_0 is as that of F, the square of its standard error σ_0^2 is more complicated. In order to calculate it we have to go back to (8) and substitute into it the new values of λ_i' obtained from the guessed weights r ,

$$\lambda_i' = \frac{N_i - 1}{(N_i - n_i)r_i^2}(\alpha' + \beta' X_i),$$

with

$$\begin{cases} \alpha' = \dfrac{N}{VS_x'^2}(S_x'^2 + \bar{x}'^2) \\[2ex] \beta' = -\dfrac{N\bar{x}'}{VS'^2}. \end{cases} \tag{20}$$

we have

$$\sigma_g^2 = \sum_{i=1}^{s} S_i^2 \frac{n_i(N_i - n_i)}{N_i - 1} \lambda_i'^2 \tag{21}$$
$$= \sum_{i=1}^{s} v_i \rho_i (\alpha' + \beta' X_i)^2,$$

where $\rho_i = S_i^2/r_i^2$. It will now be helpful to introduce notation for another kind of weighted mean and weighted (S.D.) of the X's, with weights equal to $v_i\rho_i$. So let us write

$$\bar{x}'' = \frac{\sum_i v_i \rho_i X_i}{\sum_i v_i \rho_i}; \qquad S_x''^2 = \frac{\sum_i v_i \rho_i X_i^2}{\sum_i v_i \rho_i} - \bar{x}''^2. \tag{22}$$

Expanding (21) and using (20), we have

$$\sigma_g^2 = \frac{N^2 \sum_i v_i \rho_i}{V^2} \left\{ \left[1 + \frac{\bar{x}'(\bar{x}' - \bar{x}'')}{S_x'^2} \right]^2 + \frac{\bar{x}'^2 S_x''^2}{S_x'^4} \right\}. \tag{23}$$

Formula (23) refers to the case where the X's are measured from their population, mean, $\bar{X}$. In order to reduce it to the case where the X's are given in their original values we have to substitute $(\bar{x}' - \bar{X})$ for $\bar{x}'$ and $(\bar{x}'' - \bar{X})$ for $\bar{x}''$. Thus

$$\sigma_g^2 = \frac{N^2 \sum_i v_i \rho_i}{V^2} \left\{ \left[1 + \frac{(\bar{x}' - \bar{X})(\bar{x}' - \bar{x}'')}{S_x'^2} \right]^2 + \frac{(\bar{x}' - \bar{X})^2}{S_x'^2} \frac{S_x''^2}{S_x'^2} \right\}. \tag{24}$$

Applying a similar procedure to the case where $n_1 = N_1 = 1$, but $n_i < N_i$ for $i = 2, 3, \cdots, s$, we easily find

$$F_g = N \left[y_1 - (X_1 - \bar{X}) \frac{\sum_{i=2}^{s} v_i (X_i - X_1)(y_i - y_1)}{\sum_{i=2}^{s} v_i (X_i - X_1)^2} \right], \tag{25}$$

and

$$\sigma_g^2 = N^2 (X_1 - \bar{X})^2 \frac{\sum_{i=2}^{s} v_i \rho_i (X_i - X_1)^2}{\left[\sum_{i=2}^{s} v_i (X_i - X_1)^2 \right]^2}. \tag{26}$$

This formula will help us to test the appropriateness of guesses about the values of S_i^2. It will be noticed that the λ's contain S_i^2 or r_i^2 in the same powers in the numerator and in the denominator. It follows that all we need to guess

TABLE II

Values of S_i^2

X_i	1		2		3		4		5		6		7		8		9		10		11		12		15	
Block	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2	N_i	S_i^2
1																			4	525	4	82	8	7,894		
2															2	2,052	14	4,931								
3							8	561	4	54	4	625														
4					2	46							14	1,027												
5									2	578			3	647			8	4,028								
6															4	633	12	2,031								
8													4	1,961	5	879										
9					2	61	3	71	3	2,834	2	2,294	2	586	1	961										
12	4	75	2	721			4	2,093	3	26	2	118					4	6,989							4	387
13	2	1					2	515	18	1,731	2	44														
14											7	851	16	1,380												
15											2	1,047	17	691			3	7,832								
18	2	50	2	116									2	161	6	198										
19			6	538	16	814																				
20	16	261	16	554																						
Total	24		26		20		17		30		19		58		21		41		4		4		8		4	
Weighted average		101		529		662		830		1,370		814		1,026		765		4,319		525		82		7,894		387

is a system of numbers proportional to S_i^2 and not the S_i^2 themselves. Our problem will be to test a few such guesses on the data of the Blacks Mountain Experimental Forest and see which of them gives generally a smaller value of σ_g^2

Table II gives values of the S_i^2, calculated for 15 blocks, together with the corresponding X_i. In a few cases $N_i = 1$ and consequently $S_i = 0$. These cases are not included in the table. Using the values of S_i^2 from Table II and assuming systems of the n_i's, the values of σ_F^2 were calculated for these blocks. These would be the true (S.E.)^2 of the best linear estimates of the total timber volume in each block, but it would never be possible to calculate them from sample data.

The σ_g^2's were calculated using the following guesses concerning the S_i^2: (1) That they do not depend on X_i, (2) that they are proportional to X_i, and (3) that they are proportional to $\sqrt{X_i}$. The ratios σ_F^2/σ_g^2 for all blocks taken together were found to be .770 for guess (1), .769 for guess (2), and .777 for guess (3) It is seen that, on the average, the guess that the S_i^2 are proportional to X_i gives the smallest average value of σ_g^2. It is interesting, however, to note that the differences between the three guesses are, for all practical purposes, negligible

Ratios like σ_F^2/σ_g^2 are sometimes described as the "amount of information" in F_g as compared with that in the best linear unbiased estimate F. This expression was introduced by R. A. Fisher. In certain cases it has the following property which justifies the term used: Let n be the size of the sample which serves for calculating F_g, then, if it were possible to calculate the best linear unbiased estimate F, the same accuracy of estimation would be obtained by using a smaller sample size $n\sigma_F^2/\sigma_g^2$. In the case considered in the present paper the above circumstance does not occur. Still, the ratio σ_F^2/σ_g^2 seems to be convenient to describe the situation.

7. Another scheme for estimating θ. It will be noticed that the ignorance of what are the S_i^2 is not the only circumstance which makes it difficult to apply the above formulae There is also another one connected with the values of N_i We have $N_i = 1$ in several blocks and for several different strip lengths. True this might have been avoided by defining block boundaries in such a way that $N_i \geq 2$, but it was considered best to conform strictly to the practical situation where the N_i's may be smaller. In such cases we may include in our sample all the strips of a given length, say X_1. If we apply to such samples the above formulae, deduced under the explicit assumption that the regression of Y on X is strictly linear, we shall force the fitted regression line through the point $(X_1, Y_{1\cdot})$. As the assumption of strict linearity is obviously not exact and the exhaustion of strips of length X_1 is possible only when there are very few such strips, the whole procedure may lead to serious inaccuracies in the final estimate. One safeguard against this is never to exhaust strips of any given length when dealing with formulae deduced from finite populations.

The fact that the true regression point $(X_1, Y_{1\cdot})$ does not actually lie on a

straight line makes it uncertain whether taking into account the finiteness of populations of strips of the same length is beneficial to the accuracy of the finite estimate. In the preceding sections we worked on the assumption that there is but a finite number of strips of the same length and on an inaccurate assumption that the regression is strictly linear. In the present section the first assumption will be dropped, having in mind that the effect of the inaccuracy of the second assumption may thereby be reduced.

The assumption that each of the N_i is infinite will be made only in deducing the λ_i and will be reflected in weights. Formula (11) will now reduce to $w_i = n_i/S_i^2$. If we assume further that $S_i^2 = X_i^\gamma/k$, where γ and k are some constants, then

$$\bar{w}_i = \frac{kn_i}{X_i^\gamma}; \qquad \bar{W} = \sum_i \bar{w}_i = k \sum_i \frac{n_i}{X_i^\gamma},$$

and the final estimate is

$$F = N[\bar{y} + b_0(\bar{X} - \bar{x})]. \tag{27}$$

The square of the standard error of F has again the same form as in (16),

$$\sigma_F^2 = \frac{N^2}{\bar{W}}\left[1 + \frac{(\bar{X} - \bar{x})^2}{\bar{S}_x^2}\right], \tag{28}$$

the only differences being in $\bar{W}$, $\bar{x}$, and $\bar{S}_x^2$. If $\gamma = 0$, so that the S_i^2 are assumed to be constant, then

$$\bar{w}_i = kn_i; \qquad \bar{W} = k \sum_i n_i,$$

and all the symbols $\bar{x}$, $\bar{y}$, and $\bar{S}_x^2$ assume their customary meaning of ordinary means and of ordinary $(\text{S.D.})^2$

It would be easy to deduce explicit formulae for $\gamma = 1/2$, etc., but they are not elegant and, if the necessity arises, the calculations could be carried through by starting with $\bar{w}_i = 1/X_i^\gamma$. The omission of k does not influence the form of F.

The question whether the combination of one true hypothesis about the N_i being finite, with another incorrect one that the regression is strictly linear, is better than that of two incorrect hypotheses, will be studied by means of a sampling experiment in Section 9.

8. Unbiased estimates of σ_F^2. While it may not be unreasonable to hope that a guess of a system of numbers proportional to the S_i^2 may be successful, it is entirely hopeless to try to guess the actual values of the S_i^2. It follows that, if it is desired to obtain from the sample some sort of measure of the accuracy of F, we have to calculate an estimate of σ_F^2.

We shall treat the problem by assuming that the regression of Y on X is strictly linear and that the S_i^2 are proportional to X_i^γ and that the N_i are all finite It will be noticed that they will enter the formulae by means of the

ratios $(N_i - 1)/(N_i - n_i)$. If it is desired to obtain formulae referring to the assumption of infinite N_i's, it will be sufficient to replace these ratios by unity. Of course, the symbol N will always represent the total number of strips in the actual block on which it is desired to estimate the volume of timber and will not be affected by the assumption of the N_i's being infinite

On these assumptions

$$E(y_i) = A + BX_i,$$

$$\sigma^2_{y_{ij}} = E(y_{ij} - A - BX_i)^2 = S_i^2 = kX_i^\gamma,$$

with some value of γ supposed to be accurately guessed, which however need not be specified, and with an unknown factor of proportionality, k. The square of the standard error of y_i is then known to be

$$\text{(29)} \qquad \sigma^2_{y_{i.}} = \frac{S_i^2}{n_i}\frac{N_i - n_i}{N_i - 1} = k\frac{X_i^\gamma(N_i - n_i)}{n_i(N_i - 1)}.$$

The right-hand member of (29) is equal to the reciprocal of what we have formerly denoted by w_i and described as the weight of the observations at $X = X_i$. We have mentioned above that the formula giving F does not depend on the values of the w_i, but on proportions between the w_i. In other words, if we drop the unknown factor k and denote by w_i the ratio

$$\text{(30)} \qquad \frac{n_i(N_i - 1)}{X_i^\gamma(N_i - n_i)} = w_i,$$

which involves only known quantities, these new weights will lead to exactly the same value of F as the original weights. It will now be convenient to alter the definition of weight and use formula (30). With this new meaning of w_i, (29) could be rewritten $\sigma^2_{y_{i.}} = k/w_i$.

Let us further use the letter m to denote the number of those X_i's for which we have at least one observation. In other words m will be the number of *different lengths* of strips in the sample and also the number of *different* $y_{i.}$'s that we are going to calculate from it.

Now let us go back to formula (16) giving the square of the standard error of F. We notice that, while $\bar{x}$ and S_x^2 do not depend on the unknown factor of proportionality, k, the sum W of the original weights does depend on it and with our new meaning of w_i,

$$W = \frac{1}{k}\sum_i w_i.$$

It follows that σ_F^2 should now be written in the form

$$\text{(31)} \qquad \sigma_F^2 = \frac{kN^2}{\sum_i w_i}\left[1 + \frac{(\bar{X} - \bar{x})^2}{S_x^2}\right],$$

and that, in order to estimate σ_F^2 it is sufficient to get an estimate of k. We easily get an unbiased estimate of k by merely applying the second part of the Markoff Theorem [3]. According to it an unbiased estimate of k, based on $m - 2$ degrees of freedom is given by the ratio

$$\sum_{i=1}^{m} \frac{[y_{i\cdot} - \bar{y} - b_0(X_i - \bar{x})]^2}{m-2} w_i, \tag{32}$$

in which $\bar{x}$, $\bar{y}$, and b_0 are calculated according to the assumptions made regarding N_i and γ. It may be expected, however, that the estimate (32) will not be a very accurate one because the number of degrees of freedom on which it is based may be very small.

In an attempt to find a better estimate of k we shall proceed by analogy and calculate the expectation of a sum similar to the one in the numerator of (32) but depending explicitly on the particular y_{ij}'s, namely of

$$S_0^2 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} [y_{ij} - \bar{y} - b_0(X_i - \bar{x})]^2 \frac{w_i}{n_i}.$$

It will be noticed that if the N_i are finite, y_{ij} and y_{il} are dependent and that the Theorem of Markoff does not apply to S_0^2. Introduce the notation

$$\eta_{ij} = y_{ij} - A - BX_i,$$

$$\eta_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} \eta_{ij} = y_{i\cdot} - A - BX_i.$$

Easy, but somewhat long calculations show that S_0^2 can be rewritten in the form

$$S_0^2 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{w_i}{n_i} \eta_{ij}^2 - \frac{\left(\sum_{i=1}^{m} w_i \eta_{i\cdot}\right)^2 + \frac{1}{S_x^2}\left[\sum_{i=1}^{m} w_i(X_i - \bar{x})\eta_{i\cdot}\right]^2}{\sum_{i=1}^{m} w_i},$$

which is most convenient for calculating the expectation sought. We notice first that

$$E(\eta_{ij}^2) = kX_i^\gamma,$$

$$E(\eta_{i\cdot}^2) = \sigma_{y_{i\cdot}}^2 = \frac{k}{w_i}.$$

Further, as $y_{i\cdot}$ and $y_{j\cdot}$ are mutually independent if $i \neq j$, the same is true for $\eta_{i\cdot}$ and $\eta_{j\cdot}$. It follows that

$$E(y_{i\cdot}\eta_{j\cdot}) = 0, \qquad i \neq j.$$

Consequently

$$E\left(\sum_{i=1}^{m} w_i \eta_{i\cdot}\right)^2 = E\left(\sum_{i=1}^{m} w_i^2 \eta_{i\cdot}^2\right) = \sum_{i=1}^{m} w_i^2 E(\eta_{i\cdot}^2) = k \sum_{i=1}^{m} w_i.$$

Similarly and for the same reason

$$E\left[\sum_i w_i(X_i - \bar{x})\eta_{i.}\right]^2 = kS_x^2 \sum_i w_i .$$

It follows that

$$E(S_0^2) = k\left[\sum_{i=1}^{m} \frac{n_i(N_i - 1)}{N_i - n_i} - 2\right],$$

and that the ratio

$$(33) \qquad \frac{S_0^2}{\sum_{i=1}^{m} \frac{n_i(N_i - 1)}{N_i - n_i} - 2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} [y_{ij} - \bar{y} - b_0(X_i - \bar{x})]^2 \frac{w_i}{n_i}}{\sum_{i=1}^{m} \frac{n_i(N_i - 1)}{N_i - n_i} - 2},$$

is an unbiased estimate of k. In cases where either all $n_i = 1$ or all N_i are infinite the denominator of (33) reduces to the number of degrees of freedom in S_0^2, equal to $\Sigma n_i - 2$. In other cases the denominator of (33) is greater than the number of degrees of freedom. Whether the numerical difference is large or small depends on the fractions $(N_i - 1)/(N_i - n_i)$. We may expect that in many practical cases it will be small.

We shall write

$$S_y^2 = \sum_{i=1}^{m} \frac{w_i}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \Big/ \sum_{i=1}^{m} w_i ,$$

$$r = \frac{\sum_{i=1}^{m} w_i(X_i - \bar{x})y_{i.} \Big/ \sum_{i=1}^{m} w_i}{S_x S_y}.$$

It follows that

$$S_0^2 = \sum_{i=1}^{m} w_i S_y^2(1 - r^2).$$

Substituting this formula into (33) and then the result of this substitution in place of k in (31), we finally get

$$(34) \qquad \mu_F^2 = N^2 \frac{S_y^2(1 - r^2)}{\sum_{i=1}^{m} \frac{n_i(N_i - 1)}{N_i - n_i} - 2} \left[1 + \frac{(\bar{X} - \bar{x})^2}{S_x^2}\right].$$

The case where one of the n_i is equal to the corresponding N_i, e.g., where $n_1 = N_1 = 1$ is treated in a similar manner. Using formula (18) and the notation adopted above, we can write

$$\sigma_F^2 = k \frac{N^2(X_1 - \bar{X})^2}{\sum_{i=2}^{m} w_i(X_i - X_1)^2}.$$

The unbiased estimate μ_F^2 of σ_F^2 will differ from this expression in that instead of the unknown factor k it will contain its unbiased estimate. To find this estimate we proceed exactly as above and calculate the expectation of

$$S_0^2 = \sum_{i=2}^{m} \sum_{j=1}^{n_i} [y_{ij} - y_{1\cdot} - b_0(X_i - X_1)]^2 \frac{w_i}{n_i},$$

with

$$b_0 = \frac{\sum_{i=2}^{m} w_i(X_i - X_1)(y_{i\cdot} - y_{1\cdot})}{\sum_{i=2}^{m} w_i(X_i - X_1)^2}.$$

The unbiased estimate of σ_F^2 is

$$\mu_F^2 = \frac{S_0^2}{\sum_{i=2}^{m} \frac{n_i(N_i - 1)}{N_i - n_i} - 1} \frac{N^2(X_1 - \bar{X})^2}{\sum_{i=2}^{m} w_i(X_i - X_1)^2}. \tag{35}$$

The number of degrees of freedom, f, on which μ_F^2 is based is

$$f = \sum_{i=2}^{m} n_i - 1.$$

9. Empirical tests of the preceding theory. Applications of any mathematical theory involve certain assumptions about the phenomena studied that are not exactly true. In order to have a reasonable hope that the predictions of the theory will be comparable to the actual facts, we must perform empirical tests and see whether such deviations from the assumptions underlying the theory as are *usually* met in practice influence materially or not the working of a given theory. Our object in the present section will be to test whether and to what extent such deviations influence the applicability of the theory. For that purpose it will be useful to enumerate the more important uses of the theory that are likely to be made.

The first point refers to the choice of the standard error σ_F of the best linear estimate F as the measure of accuracy with which F estimates the unknown volume of timber, θ. If all the assumptions were true, the Theorem of Liapounoff would guarantee that, when the size of the sample, Σn_i, is only moderately large, the frequency distribution of the ratio

$$(F - \theta)/\sigma_F, \tag{36}$$

would be very approximately normal about zero with unit S.E. If this were actually true then the value of σ_F would be a justifiable basis for the choice between various alternative estimates of θ. However, the discrepancies between the hypothesis underlying the theory and the actual facts may easily produce a bias in F, or may deprive σ_F of the above important property.

Therefore, the first thing that we have to test is whether in such conditions as are actually met in practice the ratio (36) is in fact distributed in repeated sampling in a way that is comparable with the normal law. The data of the 100 per cent survey of the Blacks Mountain Experimental Forest will serve us for the test.

The second important application of the theory is connected with the use of μ_F. The purpose of calculating μ_F is to characterize the accuracy of the value of F obtained from the sample. The most appropriate way of doing so is to calculate the confidence interval for θ. This has the form [5]

$$F - t_\alpha \mu_F \leq \theta \leq F + t_\alpha \mu_F \tag{37}$$

in which t_α denotes the "Student"-Fisher t taken in accordance with the number of degrees of freedom in μ_F and the chosen value of P. The confidence interval has the property that, if calculated for a great number of samples, the frequency with which the true value of θ will lie between the limits $F \pm t_\alpha \mu_F$ will approach the value $\alpha = 1 - P$ defined as the confidence coefficient.

The above statement concerning the confidence coefficient is strictly true if, apart from the various hypotheses that were enumerated, the distribution of the y's is normal. As a result of a theorem by Kozakiewicz [6] the same statement will be approximately true also for non-normally distributed y_{ij}'s, on condition that the sample sizes are considerable. In the situation where the above theory is to be applied all these assumptions are not satisfied. Still the formula for the confidence interval may well work, but before accepting this we have to have some experimental evidence. The crucial point that it must cover is whether the ratio, say

$$t = (F - \theta)/\mu_F, \tag{38}$$

does or does not follow in repeated sampling a distribution which is sufficiently close to the theoretical one, known as "Student's" distribution. If the empirical distribution of t does approach "Student's" law, then the frequency of correct statements concerning θ in the form (37) will be approximately equal to the chosen α, and conversely.

The n_i's for this experiment were fixed according to the systems shown in Table III, with all X's having a chance of appearing in the samples, and the n_i's quite closely proportional to the N_i's and approximately 25 per cent of the latter. Random sampling numbers [7] were used in making the selections of n_i strips out of any group of strips. A total of 150 block samples were drawn, equally distributed among the 15 blocks.

There are 95 samples for the case where all $n_i < N_i$. For these, formula (19) was used to calculate F and formula (24) for σ_F^2, using the guess that the S_i^2 are constant over all strip lengths. On the hypothesis that the ratio (36) is normally distributed about zero with unit S.E., we divide the range of variation of possible values of (36) into 20 intervals such that, if the hypothesis is true, then the probability of an observed value falling in any particular interval is

TABLE III

Systems of n_i's for sampling experiment

Block	X_i	N_i	n_i
1	10	4	1
	11	1	1
	12	8	2
Total		16	4
2	8	2	1
	9	14	3
Total		16	4
3	4	8	2
	5	4	1
	6	4	1
Total		16	4
4	3	2	1
	7	14	3
Total		16	4
5	5	2	1
	6	1	
	7	3	1
	9	8	2
Total		14	4
6	8	4	1
	9	12	3
Total		16	4
8	1	1	
	2	1	1
	4	1	
	6	1	
	7	4	1
	8	5	1
Total		13	3
9	3	2	1
	4	3	
	5	3	1
	6	2	1
	7	2	
	8	4	1
Total		16	4
12	1	4	1
	2	2	2
	4	4	
	5	3	1
	6	2	
	8	1	
	10	1	1
	12	1	
	14	1	
	9	4	1
	15	4	1
Total		27	7
13	1	2	
	2	1	1
	3	1	
	4	2	1
	6	2	
	5	18	4
Total		26	6
14	1	1	
	2	1	1
	3	1	
	5	1	
	6	7	2
	7	16	4
Total		27	7
15	1	1	
	2	1	
	3	1	2
	4	1	
	6	2	
	7	17	4
	9	3	1
	10	1	
Total		27	7
18	1	2	1
	2	2	
	5	1	
	6	1	1
	7	2	
	8	6	2
Total		14	4
19	2	6	1
	3	16	4
Total		22	5
20	1	16	4
	2	16	4
Total		32	8

equal to .05. For 95 samples then, the expected frequency in each interval is 4.75. The observed frequencies are shown in Table IV.

The agreement between the observed and the hypothetical distribution is tested by means of the fourth order smooth test for goodness of fit [8]. The test is designed so as to be particularly sensitive to such deviations from the hypothetical distribution that could be described as "smooth" It is used here because it is expected that, if the actual distribution of the ratios considered

TABLE IV

Frequency distribution[4] *of* $(F - \theta)/\sigma_F$ *and* $(F - \theta)/\mu_F$ *calculated under various assumptions*

	Assumption of finite population of strips				Assumption of infinite population of strips		
	$(F - \theta)/\sigma_F$		$(F - \theta)/\mu_F$		$(F - \theta)/\mu_F$		
	All $n_i < N_i$	One $n_i = N_i$	All $n_i < N_i$	One $n_i = N_i$	All $n_i < N_i$	One or more $n_i = N_i$	Total
	n_k	n_k	n_k	n_k	n_k	n_k	n_k
	5	11	4	9	3	4	7
	3	1	5	1	4	2	6
	5	2	6	2	7	2	9
	8	1	4	0	2	1	3
	3	2	4	2	5	4	9
	4	2	5	3	4	3	7
	8	1	6	2	6	0	6
	3	2	4	2	5	4	9
	3	2	5	2	8	2	10
	7	0	6	1	5	3	8
	1	0	3	0	4	0	4
	4	1	3	0	4	4	8
	6	1	5	2	7	5	12
	5	1	6	3	3	7	10
	5	1	6	1	7	1	8
	2	1	6	3	9	5	11
	5	2	8	4	6	1	7
	4	6	5	3	4	0	4
	10	1	3	2	2	6	8
	4	6	1	2	0	1	1
Total	95	44	95	44	95	55	150
ψ_4^2	1.326	33.812	5.463	13.091	9.055	2.572	8.764
P	.85	<.01	.25	.01	.06	.63	.07
$P(\chi^2)$	.57	<.01	.63	.09	.18	.12	.21

does differ markedly from the normal or from "Student's" one, then still the curve representing this actual distribution would be a "smooth" one, presumably with a single mode, and would cross the hypothetical curves at only a few points. There is empirical evidence to show [9] that in such cases the smooth test of fourth order is more powerful than the usual χ^2 test.

[4] By 20 intervals of equal probability.

The criterion used in the smooth test of the fourth order is denoted by ψ_4^2. If the hypothesis tested is true, then ψ_4^2 is distributed, approximately, as χ^2 with 4 degrees of freedom. To calculate ψ_4^2 we proceed as follows:

Let x be a random variable and H denote the hypothesis that the distribution of x is given by a perfectly specified function $f(x)$. The range of variation of x is divided into $2s = 20$ intervals,

$$(-\infty, a_1), (a_1, a_2), \cdots, (a_{19}, +\infty),$$

so that, if H is true then the probability of x falling within any such interval is exactly equal to .05. Such a subdivision can frequently be made easily from appropriate tables for $f(x)$. We associate with these intervals a variate z whose value corresponding to the kth interval will be

$$z_k = \frac{2k-1}{4s} - \frac{1}{2} = \frac{2(k-s)-1}{4s}, \qquad k = 1, 2, \cdots, 2s.$$

It will be seen that if we start at the point a_s and follow up the intervals to the right and to the left, then the corresponding values of z will be

$$z = \pm\frac{1}{4s}, \pm\frac{3}{4s}, \cdots, \pm\frac{2s-1}{4s}.$$

Consideration of the variable z is then substituted for that of the observed values $x_1, x_2, \cdots, x_n$ of x. If any value x_m falls in the kth interval $a_{k-1} < x_m \leq a_k$, then this is interpreted as an observation of x which yielded the value z_k. Let n_k denote the number of observed x's which fall in the interval (a_{k-1}, a_k) and let the Gaussian symbol $[z^i]$ stand for the sum $[z^i] = \sum_{k=1}^{2s} n_k z_k^i$. To apply the fourth order smooth test such sums have to be calculated for $i = 1, 2, 3, 4$. Then they are substituted into the equations below, deduced under the assumption that the number of intervals of subdivision of the range of x is equal to $2s = 20$.

$$u_1^2 = n^{-1}(3.468{,}440[z])^2,$$

$$u_2^2 = n^{-1}(13.500{,}884[z^2] - 1.122{,}261n)^2,$$

$$u_3^2 = n^{-1}(53.857{,}548[z^3] - 8.031{,}507[z])^2,$$

$$u_4^2 = n^{-1}(218.148{,}007[z^4] - 46.239{,}587[z^2] + 1.139{,}500n)^2.$$

Finally $\psi_4^2 = u_1^2 + u_2^2 + u_3^2 + u_4^2$. If the calculated value of ψ_4^2 exceeds the tabled value of χ^2 with four degrees of freedom, corresponding to the chosen level of significance, then the hypothesis tested, H, should be rejected.[5]

[5] The above expressions for the u's differ a little from those published in the original paper on the smooth test because in the latter the test was designed to apply only to ungrouped observations. The present formulae obtained in the Statistical Laboratory of the University of California appear in print for the first time. Obviously if the number of intervals $2s$ is increased, the formulae for grouped data will approach those for ungrouped ones.

The agreement between the observed distribution and the expected distribution is shown to be excellent in Table IV, the probability of a greater difference occurring through errors of random sampling alone being .85. The corresponding P for the χ^2 test, where consecutive pairs of intervals are combined to make 10 intervals in all, is .57.

For the case where one $n_i = N_i = 1$ there are 44 samples. For these samples the value of F was calculated from formula (25), and the value of σ_F^2 from formula (26), again taking the values of the S_i^2 as being constant over strip length within blocks. In this case the deviation from expectation shown in Table IV is greater than can be attributed to chance alone. These results are also obtained by the χ^2 test, which gives $\chi^2 = 25.091$ and $P < .01$ on 9 degrees of freedom.

The conclusions we draw from these results where one of the assumptions made is that the population of strips is finite, are that the block boundaries should be so defined that all $N_i > 1$, or if this is not done, that the systems of n_i's be such that no sampling is done from strips where $N_i = 1$. The fact that some $n_i = 0$ when the corresponding $N_i \geq 1$ has no appreciable effect on the working of the theory In the previously described test for samples in which $n_i < N_i$ the N used in formulae (19) and (24) always referred to all strips in the block, regardless of the fact that strips of some specified lengths X_i did not appear in particular samples.

Using the same samples, the distribution of $(F - \theta)/\mu_F$ is compared in a manner parallel to that described above, to the distribution of the "Student"-Fisher t, taking into account the number of degrees of freedom.

The formulae used for estimating σ_F^2, namely for calculating μ_F^2, are (34) where all $n_i < N_i$, and (35) where one $n_i = N_i$. The estimates of θ, namely F, remain unchanged from those previously calculated.

The results from this second application of the theory as judged by the smooth test in Table IV lead to the same conclusions as were made from the first application of the theory, namely, that under the assumption that the population of strips is finite no N_i should be exhausted in the sampling.

It is interesting to note that the application of the χ^2 test to the observed distribution of $(F - \theta)/\mu_F$ corresponding to samples with one $n_i = N_i = 1$, did not reject the hypothesis that it follows "Student's" law. In this case the range of t was divided into 10 intervals of equal probability and the value of χ^2 obtained was 15 091. With 9 degrees of freedom this gives P of the order of .09.

The ratio (36) cannot be determined under the assumption that the population of strips is infinite where one $n_i = N_i$ because the values of S_i^2/τ_i^2 cannot be obtained for such strips Under this assumption it is impossible to calculate σ_F by the formulae deduced in the present study and the first use of the theory must be omitted. However, the estimate of σ_F^2 from samples can be calculated and the ratio (38) determined.

The estimates F were calculated using formula (27), taking $n_i = w_i$. This same formula applies whether or not one or more of the N_i are exhausted. Each

sample from Block 15 and one sample from Block 12 exhausted two or more strip lengths and their estimates could not be calculated under the assumptions made heretofore, but these can now be obtained under the present assumptions. The estimates μ_F^2 were obtained from (34) for all samples, taking the S_i^2 as constant over all strip lengths and $n_i = w_i$. The fact that one or more N_i are exhausted does not change the procedure for such samples in any way.

For the case where all $n_i < N_i$ in Table IV, the value of $P = .06$ obtained by the ψ_4^2 test indicates that the agreement of the observed distribution with expectation, although not close, is acceptable. When the data are regrouped into 10 classes and the χ^2 test is applied, we get $P = .18$ on 9 degrees of freedom.

The ψ_4^2 test applied to the distribution of $(F - \theta)/\mu_F$ for samples where one or more $n_i = N_i$ indicates that the correspondence with expectation is good. This result is in marked contrast to the corresponding results in previous tables and bears out the belief previously expressed in Section 7, based on intuitive considerations, that by dropping the assumption of finiteness of number of strips of a given length, the error of the assumption of strict linearity of regression would be compensated for to some extent. On the basis of these findings we can add the conclusion that if, in sampling, the number of strips of a given length are exhausted, the assumption of finiteness should be dropped and the sample estimates calculated from formulae deduced under the assumption that all N_i are infinite.

There remains some question as to statistical treatment of samples in which all $n_i < N_i$, that is, whether to use formulae deduced for finite or infinite populations. The final choice can best be based on the relative size of the confidence interval (37). Where all $n_i = 1$ the estimates are the same under both assumptions. For estimates of all blocks taken together the finite population estimates tended to be within 5.5 percent of θ in 95 out of 100 trials, while the corresponding percentage for infinite population estimates was 6.0. We therefore conclude that it is better to use the assumption of finiteness of N_i where all $n_i < N_i$.

The method of sampling considered here is what could be called restricted random. The restriction consists in that we group together the sampling units of the same size, select nonrandomly several such groups, and only then proceed to draw at random n_i units of a group of N_i. Frequently the strips of the same size will be situated within the block close to one another. In those cases the restricted sampling considered will assure that the sample will contain elements more or less uniformly distributed over the area of the block.

10. Summary. Several methods of sampling timber stands and statistical treatment of the samples were considered. Data from a complete inventory of the Blacks Mountain Experimental Forest served for testing the methods in practice.

It was found that the usual method of estimating from strip samples taken within nonrectangular blocks of timber gave biased estimates, unless the linear

regression of volume on strip length passed through the origin of coordinates. It was shown that this condition was not a safe one to assume. Consequently methods of estimation were sought which were freed from this restriction.

The appropriate formulae for the best linear unbiased estimates were deduced under various combinations of the following assumptions.

(1) That the regression of timber volume on strip length is strictly linear, but may or may not pass through the origin of coordinates.

(2) That the values of the $(S.D.)^2$ of timber volumes on strips of equal lengths are (a) constant for different strip lengths, (b) proportional to strip length, and (c) proportional to the square root of strip length.

(3) That the number of strips of a given length in each block is (a) finite, and (b) infinite. Assumption (b) was based on intuitive considerations which indicated that this assumption, though known to be false, might compensate for another false assumption, namely, that of strict linearity of regression.

It was empirically found that assumption (b) of (2) gave better results than either (a) or (c). However, the advantage was small and, in the author's opinion, did not justify the extra labor in calculations which are simpler when assumption (a) is made Therefore all other calculations were made on that assumption.

An extensive sampling experiment was made to test whether the smallness of the samples combined with the conflicts between assumptions of the theory and the actual facts, influenced the validity of the normal theory.

Whenever the sample did not exhaust strips of a given length, it was found that the formulae based on the assumptions that the populations of such strips are finite and that they are infinite both work satisfactorily, generating distributions similar to those determined by the normal theory. However, the confidence intervals based on the true assumption that the populations of strips of equal length are finite, proved to be narrower. Consequently, whenever the sample does not exhaust all strips of any given length in the block, the true hypothesis concerning the number of such strips should be used. Formulae (19) and (34) are therefore the appropriate ones, using weights based on finite populations.

In cases where the sample did exhaust the strips of a given length, the treatment of the number of such strips as finite, combined with the inaccuracy of the assumption that the regression of timber volume on length of strip is linear, resulted in marked disagreement between the actual distributions of statistics and those based on normal theory. This disagreement was not found to exist in statistics calculated with formulae (27) and (34) used on the assumption of an infinity of strips of a given length. This suggests the conclusion that the exhaustion of strips of a given length by the sample should be avoided and, when this is impossible, then the formulae based on the assumption of an infinity of strips of a given length should be used.

The formulae deduced can be applied equally well to line plots as to strips. With the formulae deduced the most efficient sampling will be obtained when

the average sample strip length is close to the average for the population, but where the variation among sample strip lengths is the maximum.

The author is deeply indebted to Prof. J. Neyman for guidance and advice, and to Miss Evelyn Fix and Miss Phyllis Burleson for help in computations.

REFERENCES

[1] A. A. HASEL, *Jour. Agric. Res*, Vol. 57 (1938), p. 713.

[2] J. NEYMAN, *Lectures and Conferences on Mathematical Statistics*, Washington, D. C. (1937).

[3] F. N DAVID and J. NEYMAN, *Stat. Res. Mem.*, Vol. 2 (1938), p. 105.

[4] F N DAVID, *Stat. Res. Mem.*, Vol. 2 (1938), p. 69

[5] J. NEYMAN, *Roy Stat. Soc. Jour.*, Vol. 97 (1934), p. 558.

[6] W. KOZAKIEWICZ, *Ann. Soc. Polon Math.*, Vol. 13 (1934), p. 24.

[7] L. H. C TIPPETT, *Random Sampling Numbers*, Tracts for computers, No. 15, Cambridge University Press (1927)

[8] J. NEYMAN, *Skand. Aktuar. Tidskr.*, (1937), p. 149.

[9] J. NEYMAN, *Annals of Math. Stat*, Vol. 11 (1940), p. 478.

TABULATION OF THE PROBABILITIES FOR THE RATIO OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE TO THE VARIANCE

By B. I. Hart

Ballistic Research Laboratory, Aberdeen Proving Ground

with a note

By John von Neumann

In recent publications von Neumann has determined the distribution of δ^2/s^2, the ratio of the mean square successive difference to the variance, for odd values of the sample size n[1] and for even values of n[2] In this paper the probability function, i.e., the integral of the distribution, is evaluated for specific values of n.

Let x be a stochastic variable normally distributed with mean ζ and the standard deviation σ. The following customary definitions for the sample are:

the mean, $$\bar{x} = \frac{1}{n}\sum_{\mu=1}^{n} x_\mu,$$

the variance, $$s^2 = \frac{1}{n}\sum_{\mu=1}^{n}(x_\mu - \bar{x})^2,$$

and the mean square successive difference, $\delta^2 = \dfrac{1}{n-1}\sum_{\mu=1}^{n-1}(x_{\mu+1} - x_\mu)^2$. Letting $\dfrac{\delta^2}{s^2} = \dfrac{2n}{n-1}(1-\epsilon)$, von Neumann shows that the distribution of ϵ, $\omega(\epsilon)$, is symmetrical with zero mean and intercepts equal to $\pm\cos\dfrac{\pi}{n}$ (loc. cit [1], p. 372), and that $\omega(\epsilon)$ is determined for odd values of n by

$$\frac{d^{\frac{1}{2}(n-1)-1}}{d\epsilon^{\frac{1}{2}(n-1)-1}}\,\omega(\epsilon) = \pm\frac{(\frac{1}{2}[n-1]-1)!}{\pi}\,\frac{1}{\sqrt{\prod_{\mu=1}^{n-1}\left(\epsilon - \cos\frac{\mu\pi}{n}\right)}},$$

in the odd intervals

$$\cos\frac{\pi}{n} \geqq \epsilon \geqq \cos\frac{2\pi}{n},$$

$$\frac{\cos 3\pi}{n} \geqq \epsilon \geqq \cos\frac{4\pi}{n},\ \cdots,\ \frac{\cos(n-2)\pi}{n} \geqq \epsilon \geqq \frac{\cos(n-1)\pi}{n},$$

---

[1] John von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 367-395.

[2] John von Neumann, "A further remark on the distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat*, Vol. 13 (1942), pp 86-88

and by $\dfrac{d^{\frac{1}{2}(n-1)-1}}{d\epsilon^{\frac{1}{2}(n-1)-1}}\omega(\epsilon) = 0$ in the even intervals

$$\cos\frac{2\pi}{n} \geqq \epsilon \geqq \cos\frac{3\pi}{n},$$

$$\cos\frac{4\pi}{n} \geqq \epsilon \geqq \cos\frac{5\pi}{n}, \cdots, \cos\frac{(n-3)\pi}{n} \geqq \epsilon \geqq \cos\frac{(n-2)\pi}{n}$$

(loc. cit.[1] pp. 389–390).

For $n = 3$,

$$\omega(\epsilon) = \frac{1}{\pi}\frac{1}{\sqrt{\frac{1}{4} - \epsilon^2}} \tag{1}$$

for $\cos\frac{\pi}{3} \geqq \epsilon \geqq \cos\frac{2\pi}{3}$.

For $n = 5$,

$$\omega'(\epsilon) = \frac{1}{\pi}\frac{1}{\sqrt{-\epsilon^4 + \frac{3}{4}\epsilon^2 - \frac{1}{16}}}$$

$$\omega(\epsilon) = \frac{1}{\pi}\frac{2}{\cos\frac{\pi}{5} + \cos\frac{2\pi}{5}}\,\mathrm{sn}^{-1}\left(\left[\frac{\cos\frac{\pi}{5} + \cos\frac{2\pi}{5}}{\cos\frac{\pi}{5} - \cos\frac{2\pi}{5}}\cdot\frac{\epsilon + \cos\frac{\pi}{5}}{\cos\frac{\pi}{5} - \epsilon}\right]^{\frac{1}{2}}, \frac{\cos\frac{\pi}{5} - \cos\frac{2\pi}{5}}{\cos\frac{\pi}{5} + \cos\frac{2\pi}{5}}^{3}\right) \tag{2}$$

for $\cos\frac{\pi}{5} \geqq \epsilon \geqq \frac{2\pi}{5}$ and $\cos\frac{3\pi}{5} \geqq \epsilon \geqq \cos\frac{4\pi}{5}$.

But for $\cos\frac{2\pi}{5} \geqq \epsilon \geqq \cos\frac{3\pi}{5}$, $\omega'(\epsilon) = 0$, thus

$$\omega(\epsilon) = \text{const.} \tag{3}$$

For $n = 7$,

$$\omega''(\epsilon) = \pm\frac{2}{\pi}\frac{1}{\sqrt{-\epsilon^6 + \frac{5}{4}\epsilon^4 - \frac{3}{8}\epsilon^2 + \frac{1}{64}}} \tag{4}$$

for $\cos\frac{\pi}{7} \geqq \epsilon \geqq \cos\frac{2\pi}{7}$ and $\cos\frac{5\pi}{7} \geqq \epsilon \geqq \cos\frac{6\pi}{7}$ with the + sign, and for $\cos\frac{3\pi}{7} \geqq \epsilon \geqq \cos\frac{4\pi}{7}$ with the − sign.

But for $\cos\frac{2\pi}{7} \geqq \epsilon \geqq \cos\frac{3\pi}{7}$ and $\cos\frac{4\pi}{7} \geqq \epsilon \geqq \cos\frac{5\pi}{7}$, $\omega''(\epsilon) = 0$, thus

$$\omega'(\epsilon) = \text{const.} \tag{5}$$

---

[3] The square of the modulus. The numerical evaluation of the inverse sine amplitude function used for $n = 4, 5, 6$, is taken from unpublished tables of the Legendrian elliptic integrals by F. V Reno of the Ballistic Research Laboratory, Aberdeen Proving Ground. The square of the modulus is the argument for this tabulation.

For even values of n von Neumann shows that the distribution of ϵ,

$$\omega_{A+(0)}(\epsilon) = \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}(n-2)]\Gamma(\frac{1}{2})} \int_{\frac{|\epsilon|}{\cos\frac{\pi}{n}}}^{1} \omega_A\left(\frac{\epsilon}{\rho}\right) \rho^{\frac{1}{2}n-3}(1-\rho)^{-\frac{1}{2}}\, d\rho \qquad \text{(loc. cit.}^{2}\text{)}.$$

For $n = 4$,

$$\omega_{A+(0)}(\epsilon) = \tfrac{1}{2} \int_{\sqrt{2}\,\epsilon}^{1} \frac{\omega_A\left(\frac{\epsilon}{\rho}\right) d\rho}{\rho\sqrt{1-\rho}},$$

$$\text{where } \omega_A\left(\frac{\epsilon}{\rho}\right) = \frac{1}{\pi}\left[-\left(\frac{\epsilon}{\rho} - \cos\frac{\pi}{4}\right)\left(\frac{\epsilon}{\rho} - \cos\frac{3\pi}{4}\right)\right]^{-\frac{1}{2}}.$$

(6)
$$\omega_{A+(0)}(\epsilon) = \frac{1}{\sqrt{2}\,\pi} \int_{\sqrt{2}\,\epsilon}^{1} \frac{d\rho}{\sqrt{(\rho - \sqrt{2}\,\epsilon)(\rho + \sqrt{2}\,\epsilon)(1-\rho)}}$$

$$= \frac{\sqrt{2}}{\pi\sqrt{1+\sqrt{2}\,\epsilon}}\,\mathrm{sn}^{-1}\left(1, \frac{1-\sqrt{2}\,\epsilon}{1+\sqrt{2}\,\epsilon}^{\frac{1}{2}}\right)$$

for $\cos\frac{\pi}{4} \geqq \epsilon \geqq \frac{3\pi}{4}$.

For $n = 6$, $\omega_{A+(0)}(\epsilon) = \frac{3}{4}\int_{2\epsilon/\sqrt{3}}^{1} \frac{\omega_A\left(\frac{\epsilon}{\rho}\right)}{\sqrt{1-\rho}}\, d\rho$, where

(7) $$\omega_A\left(\frac{\epsilon}{\rho}\right) = \frac{4}{\pi(\sqrt{3}+1)}\,\mathrm{sn}^{-1}\left((\sqrt{3}+1)\left[\frac{1}{2}\cdot\frac{\sqrt{3}-2\,(\epsilon/\rho)}{\sqrt{3}+2\,(\epsilon/\rho)}\right]^{\frac{1}{2}}, 7-4\sqrt{3}^{\,\frac{1}{2}}\right)$$

for $\cos\frac{\pi}{6} \geqq \epsilon \geqq \cos\frac{\pi}{3}$ and $\cos\frac{2\pi}{3} \geqq \epsilon \geqq \cos\frac{5\pi}{6}$, and where

(8) $$\omega(\epsilon) = \text{const.}$$

for $\cos\frac{\pi}{3} \geqq \epsilon \geqq \cos\frac{2\pi}{3}$.

The integrals needed to obtain $\omega(\epsilon)$ for $n = 6$ and $\omega'(\epsilon)$ for $n = 7$ have been evaluated by numerical quadrature. Graphs of the distribution of δ^2/s^2, $\omega(\delta^2/s^2)$, for $n = 3, 4, 5, 6, 7$, are shown in Fig. 1.

The probability function, $P(\delta^2/s^2 < k) = \int_0^k \omega(\delta^2/s^2).\ d(\delta^2/s^2)$ has been obtained from $\omega(\delta^2/s^2)$ by numerical quadrature for $n = 4, 5, 6, 7$. The results are given in Table III.

As is mentioned by von Neumann, R. H. Kent has suggested a series approximation of the form

$$\omega(\epsilon) \approx \sum_{h=0}^{\infty} a_h\left(\cos^2\frac{\pi}{n} - \epsilon^2\right)^{\frac{1}{2}n-2+h},$$

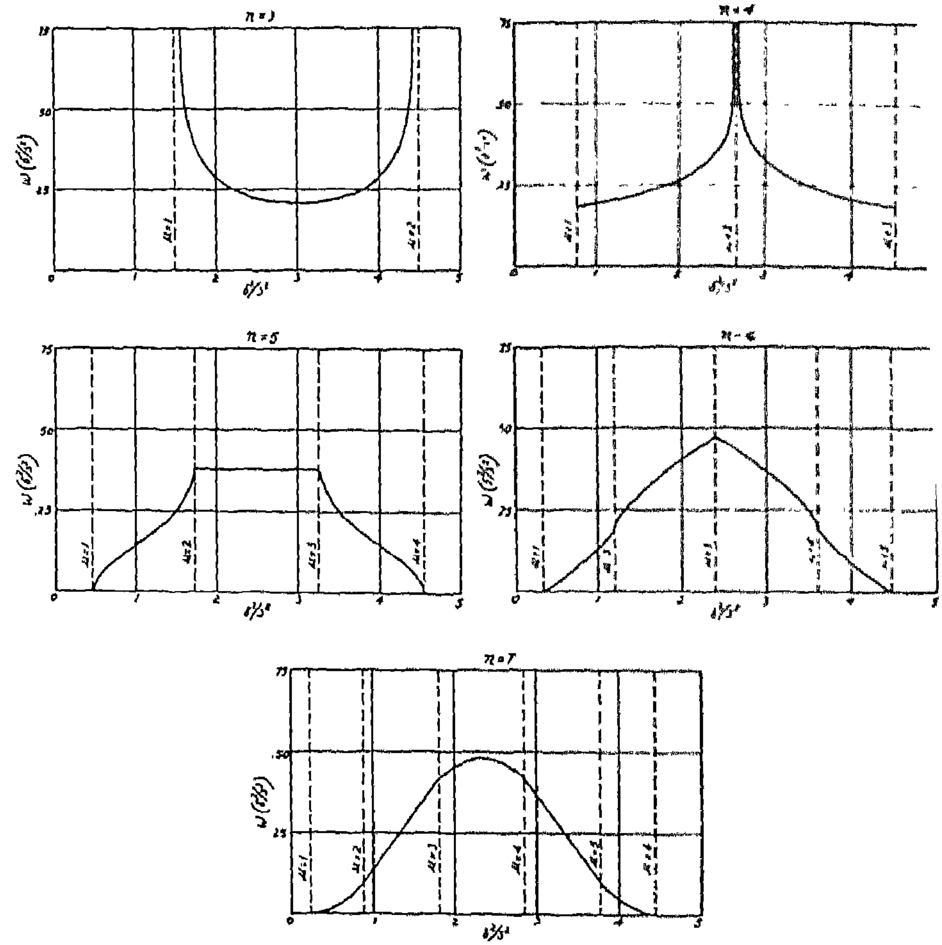


FIG. 1

since the order of vanishing of $\omega(\epsilon)$ is $\frac{1}{2}n - 2$, and since $\omega(\epsilon)$ is an even function of ϵ (loc. cit [1] p. 391). Determining the a_h by the condition of normalization and by the first three even moments of the actual distribution, M_2, M_4 and M_6 (given on pp. 377–378, loc. cit.[1]), and integrating the result, we obtain

$$P(\epsilon < k') = \int_{-\cos\frac{\pi}{n}}^{\epsilon} \sum_{h=0}^{3} a_h \left(\cos^2\frac{\pi}{n} - \epsilon^2\right)^{\frac{1}{2}n-2+h} d\epsilon$$

$$= \frac{(n-1)(n+1)(n+3)}{2^4} I_x(\tfrac{1}{2}[n-2], \tfrac{1}{2}[n-2])$$

$$\cdot \left[-\frac{1}{3} + \frac{M_2(n+5)}{\cos^2\frac{\pi}{n}} - \frac{M_4(n+5)(n+7)}{3\cos^4\frac{\pi}{n}} + \frac{M_6(n+5)(n+7)(n+9)}{45\cos^6\frac{\pi}{n}} \right]$$

$$
\begin{aligned}
&+ \frac{(n+1)(n+3)(n+5)}{2^4} I_x(\tfrac{1}{2}n, \tfrac{1}{2}n) \left[1 - \frac{M_2(3n+13)}{\cos^2 \frac{\pi}{n}} \right. \\
&\qquad \left. + \frac{M_4(3n+11)(n+7)}{3 \cos^4 \frac{\pi}{n}} - \frac{M_6(n+3)(n+7)(n+9)}{15 \cos^6 \frac{\pi}{n}} \right] \\
&+ \frac{(n+3)(n+5)(n+7)}{2^4} I_x(\tfrac{1}{2}[n+2], \tfrac{1}{2}[n+2]) \left[-1 + \frac{M_2(3n+11)}{\cos^2 \frac{\pi}{n}} \right. \\
&\qquad \left. - \frac{M_4(3n+19)(n+3)}{3 \cos^4 \frac{\pi}{n}} + \frac{M_6(n+3)(n+5)(n+9)}{15 \cos^6 \frac{\pi}{n}} \right] \\
&+ \frac{(n+5)(n+7)(n+9)}{2^4} I_x(\tfrac{1}{2}[n+4], \tfrac{1}{2}[n+4]) \left[\frac{1}{3} - \frac{M_2(n+3)}{\cos^2 \frac{\pi}{n}} \right. \\
&\qquad \left. + \frac{M_4(n+3)(n+5)}{3 \cos^4 \frac{\pi}{n}} - \frac{M_6(n+3)(n+5)(n+7)}{45 \cos^6 \frac{\pi}{n}} \right]
\end{aligned} \tag{9}
$$

The Tables of the Incomplete Beta-Function[4] can be used to evaluate (9), with $x = \frac{1}{2}\left(\frac{\epsilon}{\cos(\pi/n)} + 1\right)$. Table I shows the results obtained for the eighth and tenth moments for the distribution (9) and for the true distribution for certain values of n.

Table II gives a tabulation of $P\left(\frac{\delta^2}{s^2} < k\right)$ for $n = 7$ by the use of (9) and by the method of (4) and (5). The approximation (9) has been used for the computation of the probabilities of Table III for $n \geqq 8$[5]

It has been shown (loc. cit.[1] pp. 378–379) that for $n \to \infty$ the distribution of ϵ becomes asymptotically normal. For $n = 60$ values of δ^2/s^2 are given below for different levels of significance These values have been computed from Table III and from a table of the integral of the normal function with standard deviation equal to $\frac{2n}{n-1}\sqrt{\frac{n-2}{(n-1)(n+1)}}$, the square root of the second moment of the distribution of δ^2/s^2.

---

[4] Karl Pearson (Editor), *Tables of the Incomplete Beta-Function*, London. Biometrika Office, 1934

[5] The results obtained by L. C. Young using the Pearson Type II distribution are sufficiently precise for the significance levels and sample sizes tabulated Cf. L C. Young, "On randomness in ordered sequences," *Annals of Math. Stat*, Vol 12 (1941), pp. 293–300.

TABLE I

n	M_8 (9)	M_8 True	M_{10} (9)	M_{10} True
7	.00412	00413	.00201	00202
8	00318	00318	00150	.00151
9	.00246	.00246	00111	.00112

TABLE II

$P\left(\frac{\delta^2}{s^2} < k\right)$ for $n = 7$

k	By (9)	By (4) and (5)
.25	.00001	.00001
.30	00007	00007
.35	.00027	00027
40	00065	00065
.45	.00124	00126
.50	.00209	00214
.55	00326	.00333
60	00478	00486
65	00671	00678
70	00911	00913
.75	01203	01197
.80	01552	01534
85	01964	01932
90	02443	.02403
.95	.02995	.02957
1 00	.03624	.03598
1 05	.04333	.04325
1.10	.05126	.05137
1.15	.06006	.06036
1 20	.06976	.07020

Values of δ^2/s^2 for Different Levels of Significance

$n = 60$

	$P = .001$	$P = .005$	$P = .01$	$P = .05$
Table III.	1.2558	1.3779	1.4384	1.6082
Normal	1.2358	1.3688	1.4333	1.6092

This work was undertaken at the suggestion of Mr. R. H. Kent. I am much indebted to him and to Professor John von Neumann for many important suggestions and criticisms.

Note to Fig. 1, by John von Neumann. Inspection of the graphs of $\omega(\delta^2/s^2)$ for $n = 3, 4, 5, 6, 7$ (see Fig. 1) discloses certain singularities of the function $\omega(\delta^2/s^2)$, which seem to deserve attention.

TABLE III

$$P\left(\frac{\delta^2}{s^2} < k\right) = \int_0^k \omega\left(\frac{\delta^2}{s^2}\right) d\left(\frac{\delta^2}{s^2}\right)$$

k \ n	4	5	6	7	8	9	10	11	12
25				00001	00001	.00001	00001		
30				00007	00007	.00005	00004	00002	00001
35			00006	00027	00021	.00014	00009	.00005	00003
40			00047	00065	.00047	00031	00019	00012	00007
45			00126	00126	.00088	00059	00038	.00025	00016
.50		00038	00246	00214	00150	00103	00069	00046	00031
.55		00223	.00409	00333	.00237	00168	.00116	.00080	00055
60		00493	00615	00186	00355	00259	.00185	00132	00094
.65		00830	00865	00678	00511	00382	00282	00208	00152
70		01225	01161	00913	00710	00544	.00414	00313	00235
75		01673	01505	01197	00958	00753	.00587	00455	00351
80	00356	02171	01900	01534	01263	01015	00809	00642	00508
85	01302	02717	02348	01932	01631	01338	01089	00883	00714
90	.02257	.03310	02851	02403	.02068	01729	01436	01188	.00980
95	03223	03949	03412	.02957	02579	02196	.01858	01565	01316
1.00	04199	.04634	.04035	03598	.03171	.02745	02363	.02025	01733
1 05	.05186	05364	04728	04325	03849	.03384	02959	02578	.02241
1 10	06184	.06140	05500	05137	04618	.04120	03655	03232	02852
1 15	07194	.06963	.06361	06036	.05482	04957	04458	03997	03577
1.20			07323	07020	06445	05901	05375	04882	04425
1 25						.06956	06412	05894	05407
1 30								07040	06531

k \ n	15	20	25	30	40	50	60
35	00001						
40	00002						
45	00004						
.50	.00009	.00001					
.55	.00018	00002					
60	.00033	00005	.00001				
65	00059	.00012	00002				
70	00100	.00024	.00005	.00001			
75	00161	.00044	00011	00003			
80	00250	00076	.00023	.00007	00001		
.85	.00375	.00127	00044	00015	00002		
90	.00547	00206	00079	00030	00004	.00001	
95	00778	.00323	00135	00057	.00010	00002	
1 00	.01079	.00489	00222	00102	.00022	00005	00001
1 05	01465	00720	00355	.00176	00044	00012	.00003
1 10	.01950	01033	00550	00294	00085	00026	.00008
1 15	02550	01448	00826	00474	00158	00054	'00019
1 20	03280	01986	01208	.00738	.00280	00108	00043
1 25	04155	02670	01723	01117	00476	00206	00092
1 30	05189	03524	02402	.01644	00780	.00376	00185
1.35	06396	04571	03276	02357	01235	.00656	00355
1.40	07787	05834	04379	03298	01892	.01098	.00649
1 45		.07333	.05743	04511	02810	01769	.01133
1 50			07398	06038	.04055	02750	01893
1.55				07920	05696	.04131	.03034
1.60					07797	.06006	.04075
1.65						.08465	.06912
1 70							.09949

Values of k for which $P\left(\frac{\delta^2}{s^2} < k\right) = 0$

n	k	n	k
4	7811	15	0468
5	4775	20	.0259
6	.3215	25	0164
7	2311	30	0113
8	1740	40	.0063
9	.1357	50	0040
10	.1088	60	0028
11	0891		
12	.0743		

Furthermore it may be shown that

$$(2) \qquad F(x) = \int_{-\infty}^{x} F'(S)\, dS, \qquad F'(x) = f(x),$$

where $f(x)$ is the ordinary probability function. Also

$$(3) \qquad P(a \leq x \leq b) = \int_{a}^{b} f(x)\, dx.$$

Similarly for the discrete case,

$$(4) \qquad F(x) = \sum_{i<x} {}_h f(i), \qquad \underset{h}{\Delta} F(x) = \frac{1}{h} f(i),$$

$$(5) \qquad P(a \leq x \leq b) = F(b + h) - F(a) = \sum_{i=a}^{b} {}_h f(i),$$

where a, b are among the values $nh + d$, and $\underset{h}{\Delta}$ is the usual h-difference. In both cases the percentiles are given by the solutions of the equations

$$(6) \qquad F(x) = n/100.$$

Equations (1), (3) and (5) formulate the advantage to the direct use of $F(x)$, which was mentioned in section 1. Related to this is the fact that the process of finding $f(x)$ from $F(x)$ is at least theoretically much simpler than conversely, as (2) and (4) show. The directness of equation (6) is often an advantage also.

The main problems confronting one in trying to utilize these advantages are (a) to find suitable cumulative functions and (b) to find methods of fitting $F(x)$ directly. These are next discussed.

3. Some special functions $F(x)$. An obvious method of attack is to use (2) or (4) on some $f(x)$. The integration involved is precisely the difficulty the writer wishes to avoid. The cumulative function might be sought directly in probability theory. A differential equation incorporating some of the properties of $F(x)$ given in section 2 is

$$(7) \qquad \frac{dy}{dx} = y(1 - y)g(x, y), \qquad y = F(x),$$

where $g(x, y)$ is to be positive for $0 \leq y \leq 1$ and x in the range over which the solution is to be used. It is to be noted that (7) is very similar to the differential equation

$$\frac{dy}{dx} = y(m - x)g(x, y), \qquad y = f(x),$$

which generates the Pearson system if $g(x, y) = (a + bx + cx^2)^{-1}$

Equation (7) implies the non-decreasing property for $F(x)$, while for many choices of $g(x, y)$, dy/dx will be zero at $y = 0$ and $y = 1$. When $g(x, y) = g(x)$, (7) becomes

$$F(x) = [e^{-\int g(x)\,dx} + 1]^{-1}. \tag{8}$$

Some functions $g(x)$ whose integrals are such that $F(x)$ increases from 0 to 1 on the interval $-\infty < x < \infty$ are c, cx^{-1}, $[(c - x)x]^{-1}$, $c \sec^2 x$ and $c \cosh x$, where $c > 0$ Generalizations of their corresponding $F(x)$ are given below in (10)–(14) respectively.

Another method of attack is to simply consider functions which have the properties given in section 2. The assumption of high contact provides for the existence of certain integrals to be discussed in section 5. Many functions having the required properties are to be found in tables of definite integrals, particularly Bierens de Haan [1].

A list of particular $F(x)$ is given below. In all cases the number of parameters would be increased by two by letting $x = \gamma x' + \delta$, where γ and δ fix the origin and scale. These parameters are determined by $\bar{x}$ and σ The range of x over which the given expression is to be used is written to the right when it is not $(-\infty, \infty)$. Constants k, r and c are positive real numbers.

$$F(x) = x, \qquad (0, 1), \tag{9}$$

$$F(x) = (e^{-x} + 1)^{-r}, \tag{10}$$

$$F(x) = (x^{-k} + 1)^{-r}, \qquad (0, \infty), \tag{11}$$

$$F(x) = \left[\left(\frac{c - x}{x}\right)^{1/c} + 1\right]^{-r}, \qquad (0, c), \tag{12}$$

$$F(x) = (ke^{-\tan x} + 1)^{-r}, \qquad \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \tag{13}$$

$$F(x) = (ke^{-c \sinh x} + 1)^{-r}, \tag{14}$$

$$F(x) = 2^{-r}(1 + \tanh x)^r, \tag{15}$$

$$F(x) = \left(\frac{2}{\pi} \arc\tan e^x\right)^r, \tag{16}$$

$$F(x) = 1 - \frac{2}{k[(1 + e^x)^r - 1] + 2}, \tag{17}$$

$$F(x) = (1 - e^{-x^2})^r, \qquad (0, \infty), \tag{18}$$

$$F(x) = \left(x - \frac{1}{2\pi} \sin 2\pi x\right)^r, \qquad (0, 1), \tag{19}$$

$$F(x) = 1 - (1 + x^c)^{-k}, \qquad (0, \infty), \tag{20}$$

Most of these functions have unimodal probability functions $f(x)$, and all of the functions may be readily handled from the calculational standpoint. To

check upon their suitability for practical work, the values of α_3 and α_4 for some special cases were obtained approximately by evaluating $F(x)$ at a convenient regular interval, differencing, and using the results as frequencies of a discrete

TABLE I

Calculated α_3 and α_4 for special functions $F(x)$

Function	Parameters	α_3	α_4
(15)	$r = 1$	0	4.01
(16)	$r = 1$	0	3.24
(17)	$k = 1, r = 2$	$-.62$	4.50
(17)	$k = 2, r = 1$	0	4.11
(17)	$k = 2, r = 2$	$-.51$	4.22
(18)[2]	$r = 1$	.63	3.25
(19)[2]	$r = 1$	0	2.41

variable. No correction for grouping was made. The values of α_3 and α_4 for several of the above functions are given in Table I, where

$$\mu_j' = \int_{-\infty}^{\infty} x^j f(x)\, dx, \qquad \sum_{i=-\infty}^{\infty} i^j f(i)$$

$$\mu_j = \int_{-\infty}^{\infty} (x - \mu_1')^j f(x)\, dx, \qquad \sum_{i=-\infty}^{\infty} (i - \mu_1')^j f(i) \tag{21}$$

$$\alpha_j = \frac{\mu_j}{\sigma^j}, \qquad \sigma^2 = \mu_2 .$$

It will be seen that a variety of values of α_4 appear. The values of α_3 vary considerably in most cases as r varies. These functions show promise of being useful after further investigation. The values of α_3 and α_4 for (20) are convenient and adaptable. This function will be discussed in detail in section 6.

4. Methods of fitting $F(x)$. The problem of graduation of data by a cumulative function involves three steps. (a) the selection of the type of function (b) the determination of the parameters of the function, and (c) the graduation. The first two are often determined by such moment characteristics as α_3 and α_4, as in the Pearson system of frequency functions. The third step involves integration or summation if $f(x)$ is used, whereas, once $F(x)$ is fitted, all that remains to be done is evaluation of the function and differencing.

To fit $F(x)$ by moments, it must be possible to determine the parameters of $F(x)$ from $\bar{x}$, σ, α_3 and α_4. The cumulative moments described in the next section, when they can be evaluated, will lead to the values of the $\bar{x}$, σ, α_3 and α_4 for various values of the parameters. If the relations between the parameters and the moments are difficult or impossible to obtain, then tables may be constructed and interpolation used. The usual process would be to use the α_3

[2] The method of moments of section 5 was used for these values.

and α_4 tables to determine the primary parameters such as c, k and r in (9)–(20). Then for the given values of c, k, r, one computes the corresponding values of $\bar{x}$ and σ from their tables, and these are used to obtain the parameters γ and δ for $x = \gamma x' + \delta$. This procedure is illustrated in section 6.

Even when the cumulative moments cannot be evaluated, this method is still possible. Graduation by a small interval is used to construct tables of $\bar{x}$, σ, α_3 and α_4 for varying values of the parameters. Then the table can be used as described above Thus it is seen that in practice any $F(x)$ can be fitted by this technique

The usefulness of a cumulative or a probability function depends upon how wide a range of sets of values of the α_i the function covers, and whether such values occur in practice In most of the functions (9)–(20), α_3 and α_4 are continuous functions of the parameters. If there is only one parameter then only α_3 (or α_4) can be fitted in the range of values of α_3 which the function possesses, but in the case of two parameters both α_3 and α_4 can be fitted. Three or more parameters permit α_5 etc. to be fitted

5. Cumulative moment theory for $F(x)$. A moment definition for $F(x)$ is now presented Since for $n \geq 0$, $\lim_{b \to \infty} \int_a^b x^n F(x)\,dx = \infty$, $\int_{-\infty}^{\infty} x^n F(x)\,dx$ cannot be used However, it was assumed in section 2 that for some $k > j + 1$, $[1 - F(x)]x^k$ is ultimately bounded Hence, $\lim_{x \to \infty} [1 - F(x)]x^j = 0$. Thus $1 - F(x)$ can be used as a factor when integrating over any interval (a, ∞), a being finite. But the factor $F(x)$ must be used for an interval of the type $(-\infty, b)$ Two integrals are needed, and we define the *cumulative moment*, $M_j(a)$, by

$$(22) \qquad M_j(a) = \int_a^{\infty} (x - a)^j [1 - F(x)]\,dx - \int_{-\infty}^{a} (x - a)^j F(x)\,dx,$$

which exists under the assumptions of section 2 The difference of the integrals is used because, as will be shown, this leads to simpler results than could be obtained by addition If $a = 0$ in (22) then calling $M_j(0) = M_j$,

$$(23) \qquad M_j = \int_0^{\infty} x^j [1 - F(x)]\,dx - \int_{-\infty}^{0} x^j F(x)\,dx.$$

Definitions for the discrete case are similar

$$(24) \qquad M_j(a) = h \sum_{i=a+h}^{\infty}{}_h\, (i - a)^{(j)h}[1 - F(i)] - h \sum_{i=-\infty}^{a}{}_h\, (i - a)^{(j)h} F(i),$$

$$(25) \qquad M_j = h \sum_{i=h}^{\infty}{}_h\, i^{(j)h}[1 - F(i)] - h \sum_{i=-\infty}^{0}{}_h\, i^{(j)h} F(i),$$

where $i^{(j)h} = i(i - h) \cdots (i - \overline{j - 1}h)$. This function is used because it has simpler properties in the finite calculus than has i^j.

Various relations between the cumulative moments $M_j(a)$ and M_j, and between these and μ_j', μ_j and α_j of (21) are now developed. To express $M_j(a)$ in terms of M_i's, use $(x-a)^j = \sum_{i=0}^{j} {}_jC_i x^{j-i}(-a)^i$. Thus,

$$M_j(a) = \int_a^\infty (x-a)^j[1-F(x)]\,dx - \int_{-\infty}^a (x-a)^j F(x)\,dx$$

$$= \int_0^\infty (x-a)^j[1-F(x)]\,dx - \int_{-\infty}^0 (x-a)^j F(x)\,dx - \int_0^a (x-a)^j\,dx$$

$$M_j(a) = \sum_{i=0}^{j} {}_jC_i(-a)^i M_{j-i} + \frac{(-a)^{j+1}}{j+1}. \tag{26}$$

One reason for the minus sign of (22) may be noted here, because in the contrary case the last term would be $\int_0^a (x-a)^j[2F(x)-1]\,dx$. By translating the origin in (26) to $x = a$, renaming the moments, and replacing $-a$ by a, one obtains

$$M_j = \sum_{i=0}^{j} {}_jC_i a^i M_{j-i}(a) + \frac{a^{j+1}}{j+1}. \tag{27}$$

To bring in ordinary moments, integration-by-parts and (2) are used.

$$M_j(a) = \left[\frac{(x-a)^{j+1}}{j+1}\{1-F(x)\}\right]_a^\infty + \int_a^\infty \frac{(x-a)^{j+1}}{j+1} f(x)\,dx$$

$$- \left[\frac{(x-a)^{j+1}}{j+1} F(x)\right]_{-\infty}^a + \int_{-\infty}^a \frac{(x-a)^{j+1}}{j+1} f(x)\,dx \tag{28}$$

$$= \frac{1}{j+1}\int_{-\infty}^\infty (x-a)^{j+1} f(x)\,dx,$$

the first and third quantities vanishing because of the contact assumption. A second justification of the minus sign of (22) appears here, since if a positive sign were used, the fourth term would have been subtracted and the integrals would not combine into (28). Expansion of $(x-a)^{j+1}$ in powers of x and $x - \mu_1'$ yields respectively

$$M_j(a) = \frac{1}{j+1}\sum_{i=0}^{j+1} {}_{j+1}C_i(-a)^i \mu_{j+1-i}', \tag{29}$$

$$M_j(a) = \frac{1}{j+1}\sum_{i=0}^{j+1} {}_{j+1}C_i(\mu_1'-a)^i \mu_{j+1-i}. \tag{30}$$

Also setting $a = 0$,

$$M_j = \frac{1}{j+1}\mu_{j+1}' \tag{31}$$

$$M_j = \frac{1}{j+1}\sum_{i=0}^{j+1} {}_{j+1}C_i \mu_1'^i \mu_{j+1-i}. \tag{32}$$

It may be shown that the existence of $M_j(a)$ implies that of the μ_i' $i = 1, \cdots, j+1$, and conversely if μ_j' exists then so do the $M_i(a)$ $i = 0, \cdots, j-1$ The following formulas are obtained by the opposite integration by parts, taking two different forms for $\int f(x)\,dx$: $F(x)$ and $-[1-F(x)]$, to avoid indeterminate situations.

$$\mu_j' = \int_a^\infty x^j f(x)\,dx + \int_{-\infty}^a x^j f(x)\,dx$$

$$= -[x^j\{1 - F(x)\}]_a^\infty$$

$$+ j\int_a^\infty x^{j-1}[1 - F(x)]\,dx + [x^j F(x)]_{-\infty}^a - j\int_{-\infty}^a x^{j-1}F(x)\,dx.$$

The first and third terms vanish by the contact assumption. Then using $(x - a + a)^{j-1}$ for x^{j-1},

$$\mu_j' = j\sum_{i=0}^{j-1} {}_{j-1}C_i\, a^i M_{j-1-i}(a) + a^j, \qquad j > 0 \tag{33}$$

Also in the same manner

$$\mu_j = j\sum_{i=0}^{j-1} {}_{j-1}C_i(a - \mu_1')^i M_{j-1-i}(a) + (a - \mu_1')^j, \qquad j > 1,$$

or

$$\mu_j = j\sum_{i=0}^{j-1} {}_{j-1}C_i[-M_0(a)]^i M_{j-1-i}(a) + [-M_0(a)]^j, \qquad j > 1, \tag{34}$$

using (29) $M_0(a) = \mu_1' - a$. Letting $a = 0$,

$$\mu_j' = jM_{j-1}, \qquad j > 0 \tag{35}$$

$$\mu_j = j\sum_{i=0}^{j-1} {}_{j-1}C_i(-M_0)^i M_{j-1-i} + (-M_0)^j, \qquad j > 1. \tag{36}$$

An interesting graphical property of $F(x)$ may be seen from (35) $j = 1$ by taking $\mu_1' = 0$. Then $M_0 = 0$ and hence $\int_0^\infty [1 - F(x)]\,dx = \int_{-\infty}^0 F(x)\,dx$. Thus the mean is that ordinate which equates the two areas bounded by (i) $y = F(x)$, $y = 0$ and $x = \mu_1'$ and (ii) $y = F(x)$, $y = 1$ and $x = \mu_1'$.

It is worth noting that the expressions (34) and (36) have the same coefficients, independent of a. This is to be expected because of the invariance of μ_j under translation.

If $a = \mu_1'$ then (30) simplifies to $M_j(\mu_1') = \frac{1}{j+1}\mu_{j+1}$. Lastly, expressions for α_j's in terms of the $M_i(a)$'s are given.

$$\alpha_3 = \frac{3M_2(a) - 6M_1(a)M_0(a) + 2M_0^3(a)}{[2M_1(a) - M_0^2(a)]^{3/2}}$$

$$(37)\qquad \alpha_4 = \frac{4M_3(a) - 12M_2(a)M_0(a) + 12M_1(a)M_0^2(a) - 3M_0^4(a)}{[2M_1(a) - M_0^2(a)]^2}$$

$$\alpha_j = \frac{j\sum_{i=0}^{j-1} {}_{j-1}C_i[-M_0(a)]^i M_{j-1-i}(a) + [-M_0(a)]^j}{[2M_1(a) - M_0^2(a)]^{j/2}} .$$

The discrete case has been carried through in an exactly similar manner, by the use of finite rather than infinitesimal calculus. Only the results will be stated here The notation used is that of Steffensen [2].

$$(38)\qquad M_j(a) = \sum_{r=0}^{j} {}_jC_{j-r}[a + (r - 1)h]^{(r)h}(-1)^r M_{j-r} + \frac{(-1)^{j+1}}{j+1}[a + (j - 1)h]^{(j+1)h}, \qquad j > 0$$

$$(39)\qquad M_0(a) = M_0 + a$$

$$(40)\qquad M_j = \sum_{r=0}^{j} {}_jC_r\, a^{(r)h} M_{j-r}(a) + \frac{(a + h)^{(j+1)h}}{j + 1}$$

$$(41)\qquad M_j(a) = \frac{1}{j+1}\sum_{r=0}^{j+1}\mu_r'\sum_{k=r}^{j+1} {}_kC_r\frac{D^k 0^{(-j-1)}}{k!}(-h)^{j+1-k}(h - a)^{k-r}, \qquad j > 0$$

$$(42)\qquad M_0(a) = \mu_1' - a$$

$$(43)\qquad M_j(a) = \frac{1}{j+1}\sum_{r=0}^{j+1}\mu_r\sum_{k=r}^{j+1} {}_kC_r\frac{D^k 0^{(-j-1)}}{k!}(-h)^{j+1-k}(\mu_1' + h - a)^{k-r}, \qquad j > 0$$

$$(44)\qquad M_j = \frac{1}{j+1}\sum_{r=0}^{j+1}\mu_r' h^{j+1-r}\sum_{k=r}^{j+1} {}_kC_r\frac{D^k 0^{(-j-1)}}{k!}(-1)^{k+j+1}, \qquad j > 0$$

$$(45)\qquad M_0 = \mu_1'$$

$$(46)\qquad M_j = \frac{1}{j+1}\sum_{r=0}^{j+1}\mu_r\sum_{k=r}^{j+1} {}_kC_r\frac{D^k 0^{(-j-1)}}{k!}(-h)^{j+1-k}(\mu_1' + h)^{k-r}, \qquad j > 0$$

$$(47)\qquad \mu_j' = a^j + \sum_{r=0}^{j-1} M_r(a)\sum_{k=r+1}^{j} h^{j-k}\frac{\Delta^k 0^j}{k!}\,{}_kC_r(k - r)(a - h)^{(k-r-1)h}$$

$$(48)\qquad \mu_j = [-M_0(a)]^j + \sum_{r=0}^{j-1} M_r(a)\sum_{k=r+1}^{j} h^{j-k}\frac{\Delta^k 0^j}{k!}\,{}_kC_r(k - r)[-M_0(a) - h]^{(k-r-1)h}$$

$$(49)\qquad \mu_j' = \sum_{r=0}^{j-1} M_r h^{j-r-1}\sum_{k=r+1}^{j}\frac{\Delta^k 0^j}{k!}\frac{k!}{r!}(-1)^{k-r-1}$$

$$(50)\qquad \mu_j = (-M_0)^j + \sum_{r=0}^{j-1} M_r\sum_{k=r+1}^{j} h^{j-k}\frac{\Delta^k 0^j}{k!}\,{}_kC_r(k - r)[-M_0 - h]^{(k-r-1)h}.$$

The writer has verified that under certain fairly general conditions the discrete case (38)–(50) approaches the continuous case (26)–(36) as $h \to 0$

The following three propositions are merely stated without proof since they follow so immediately from (23), (25), (31), (45), (21), (2) and (4).

PROPOSITION 1: *Given a set of functions* $F_i(x)$ *and positive constants* k_i $i = 1, \cdots, n$ *for which* $\sum_{i=1}^{r} k_i = 1$, *then for* $F(x) = \sum_{i=1}^{r} k_i F_i(x)$, $M_j = \sum_{i=1}^{r} k_{i\,i} M_j$ *if all the latter exist.*

PROPOSITION 2. *In the above notation, if all the* ${}_i\mu_1'$ *are equal, then* $\mu_j = \sum_{i=1}^{r} k_{i\,i}\mu_j$, *when the latter exist*

PROPOSITION 3. *If in addition to the above hypotheses, all the* ${}_i\mu_2$ *are equal, then*

$$\alpha_j = \sum_{i=1}^{r} k_{i\,i}\alpha_j . \tag{51}$$

These propositions are sometimes convenient in forming a linear combination of functions $F(x)$, to obtain a function with desired properties. It may be noted that Proposition 1 is still algebraically true even with negative k_i's, but these might give negative derivatives $f(x)$ for $F(x)$

6. An algebraic function, $F(x) = 1 - \dfrac{1}{(1 + x^c)^k}$. This simple algebraic cumulative function will be discussed in detail. The α_i can be calculated directly by the application of (23), (36) and (21). The resulting α_3 and α_4 values cover a broad range, within which those of many empirical and theoretical distributions lie. A method of finding such cumulative functions with desired α_3 and α_4 will be given. Several graduations are presented for illustration.

This function appears in Bierens de Haan [1] and has the desired properties. The writer has not yet found a probability justification for the function. However, since the α_i are so close to those of functions which can be so supported, it seems that it may eventually prove to be at least some definite approximation to a probability situation

The complete definition is

$$F(x) = 1 - \frac{1}{(1 + x^c)^k} \qquad x \geq 0 \tag{52}$$
$$= 0 \qquad x < 0,$$

where c, $k \geq 1$ are real numbers. The probability function

$$F'(x) = f(x) = \frac{kcx^{c-1}}{(1 + x^c)^{k+1}}, \tag{53}$$

is unimodal at $x = \left(\dfrac{c - 1}{ck + 1}\right)^{1/c}$ if $c > 1$, and L-shaped if $c = 1$.

Use of (23) on (52) gives

$$M_j = \int_0^\infty \frac{x^j\,dx}{(1+x^c)^k}, \qquad j < ck - 1. \tag{54}$$

But from Bierens de Haan [1]

$$\int_0^\infty \frac{x^{g-1}\,dx}{(1+x^c)^k} = \frac{(c-g)^{[k-1]_c}\,\pi}{c^k(k-1)!\sin(g\pi/c)} \qquad g < c, \tag{55}$$

where $a^{[r]_c} = a(a+c)\cdots(a+\overline{r-1}c)$.
Hence

$$M_j = \frac{(c-j-1)^{[k-1]_c}\,\pi}{c^k(k-1)!\sin\dfrac{j+1}{c}\pi}, \qquad j < c - 1. \tag{56}$$

However, if $j \geq c - 1$ then (55) can still be used through reducing the exponent of x by $x^{j-c}(1+x^c) - x^{j-c} = x^j$. (56) is only good for integral values of k. A more general formula is obtainable by letting $(1 + x^c) = 1/s$. Then

$$M_j = \frac{1}{c}\int_0^1 (1-s)^{(j+1)/c-1} s^{k-(j+1)/c-1}\,ds$$

$$= \frac{1}{c}B\left(\frac{j+1}{c}, k - \frac{j+1}{c}\right)$$

$$M_j = \frac{\Gamma\left(\dfrac{j+1}{c}\right)\Gamma\left(k - \dfrac{j+1}{c}\right)}{c\Gamma(k)}, \tag{57}$$

for $j = 0, 1, \cdots$ up through $j < ck - 1$, and c, k any real numbers ≥ 1. To determine the μ_j values the easiest way is to compute the values of the M_j by (56) or (57), and then to use (36):

$$\mu_2 = 2M_1 - M_0^2, \qquad \mu_3 = 3M_2 - 6M_1M_0 + 2M_0^3,$$

$$\mu_4 = 4M_3 - 12M_2M_0 + 12M_1M_0^2 - 3M_0^4, \text{ etc.}$$

Having these, definitions (21) are used for the α_j.

The results for some integral values of k and c are given in Tables II and III. These computations were made from (56). Formula (57) shows that for a fixed c, M_j for $k + 1$ is obtained by multiplying M_j for k by $\dfrac{kc - j - 1}{kc}$. This recursion relation is very helpful in the computation, because it enables all of the values of the M_j's for a given c to be found from those for the lowest value of k for which M_j exists. The values which need to be copied down in the computation for μ_1', σ, α_3, α_4, by a calculating machine are M_0, M_1, M_2, M_3, M_0^2, M_0^3, M_0^4, $6M_0$, $12M_0$, $12M_0^2$, μ_2, σ, σ^3, σ^4, μ_3, α_3, μ_4, α_4. Because of heavy cancellation, especially in μ_3 and μ_4, it seemed advisable to use eight signi-

ficant figures throughout. Eight-place sines were obtained from Gifford [3]. The values of the M_j for $k = 11$ were also checked by eight-place logarithms [4]. These verify the values of the M_j for $k < 11$ because of the recurrence calculation.

TABLE II

Mean μ_1', and Standard Deviation σ for $F(x) = 1 - \dfrac{1}{(1 + x^c)^k}$

(In each cell the upper number is μ_1' and the lower number is σ)

k \ c	1	2	3	4	5	6	7	8	9	10
1	— —	1.57080 —	1 20920 97787	1 11072 .58060	1.06896 42265$^-$	1.04720 .33552	1.03438 27953	1.02617 24019	1.02060 21089	1 01664 18815
2	1.00000 —	.78540 .61899	.80613 .39533	.83304 30239	.85517 24794	87266 .21116	.88661 18433	89790 16375	.90720 .14742	.91498 .13411
3	.500000 86603	.58905$^-$ .39118	.67178 29349	.72891 .24029	.76965$^-$ .20461	.79994 17852	.82328 .15847	.84178 14253	.85680 .12953	86923 11872
4	.33333 .47140	49087 .30393	.59714 .24784	.66817 21077	.71834 18344	.75550 16234	78408 .14555$^-$	.80671 13187	.82507 12053	.84025 11097
5	25000 .32275	.42951 .25596	.54737 .22070	.62641 .19269	.68242 .17028	72402 .15220	.75607 .13743	.78150$^-$ .12517	.80215$^-$ .11487	.81925$^-$ .10610
6	20000 24495$^-$	.38656 .22488	51088 20220	59509 .18010	.65513 .16103	.69980 .14505$^-$	.73447 .13168	76196 .12043	78432 .11087	80286 .10266
7	.16667 19720	.35435$^-$ .20274	.48250 .18851	.57029 .17064	.63329 15403	.68045$^-$ .13962	71698 .12731	.74609 .11682	.76980 .10782	.78948 10004
8	.14286 .16496	32904 .18599	.45952 17783	.54992 .16316	.61520 14846	.66425$^-$ 13528	.70235 .12383	73276 11394	.75758 .10539	.77820 .09796
9	.12500 14174	30847 .17275	44038 16918	.53274 15704	.59982 14389	65041 .13171	.68981 12095$^-$	.72131 .11156	.74706 .10338	.76848 .09623
10	.11111 .12423	.29134 16197	.42407 .16197	.51794 .15190	58649 .14002	.63836 .12869	.67886 .11851	.71130 .10954	.73783 .10168	75994 .09477
11	.10000 .11055	27677 .15297	.40993 .15584	50499 .14749	.57476 .13669	62772 .12608	.66916 .11640	.70240 10780	.72964 .10021	.75231 .09351

It will be seen from Table II that in most cases the values of α_3 and α_4 lie within useful ranges. The graph shows the general relationship between α_3, k and c. The curves are the traces of planes $k = 1, 2, \cdots$ upon the surface $\alpha_3 = G(c, k)$. Other traces would contain all pairs (c, k) giving a fixed α_3.

TABLE III

Skewness α_3 and Kurtosis α_4 for $F(x) = 1 - \dfrac{1}{(1 + x^c)^k}$

(In each cell the upper number is α_3 and the lower number is α_4)

k \ c	1	2	3	4	5	6	7	8	9	10
1	— —	— —	— —	4.285^{-} —	2.485 20.56	1.820 14.77	1.458 10.36	1.225 8.342	1.060 7.215	.937 6.510
2	— —	4.086 —	1.589 10.81	.956 5.037	.635 4.630	.134 4.106	.294 3.859	.190 3.736	.109 3.673	044 3.646
3	— —	1.909 12.46	.919 5.132	.513 3.871	.277 3.485	.119 3.358	$.005^{-}$ 3.329	−.083 3.343	−.152 3.376	−.208 3.418
4	7.071 —	1.432 7.356	.682 4.036	$.335^{-}$ 3.303	$.125^{-}$ 3.189	−.019 3.169	−.125 3.205	−.207 3.263	−.271 3.327	−.325 3.393
5	4.648 73.80	1.218 5.832	559 3.604	.238 3.154	.040 3.070	−.097 3.098	−.199 3.165	−.277 3.243	−.340 3.324	−.391 3.401
6	3.810 38.67	1.094 5.118	.484 3.380	.178 3.045	−.013 3.010	−.147 3.005	−.216 3.150^{-}	−.323 3.241	−.384 3.330	−.435 3.416
7	3.381 27.86	1.014 4.707	433 3.245^{-}	.136 2.979	−.051 2.975	−.181 3.048	−.279 3.144	$-.355^{-}$ 3.244	−.415 3.339	−.465 3.430
8	3.118 22.73	.958 4.443	.396 3.154	.106 2.936	−.078 2.953	−.207 3.039	−.303 3.143	−.378 3.248	−.438 3.349	−.488 3.442
9	2.940 19.76	916 4.258	.368 3.091	.083 2.906	−.098 2.938	−.226 3.033	−.322 3.143	−.396 3.252	−.456 3.357	$-.505^{-}$ 3.453
10	2.811 17.83	.884 4.122	347 3.043	065 2.883	$-.115^{-}$ 2.928	−.242 3.030	−.336 3.144	−.410 3.257	−.470 3.364	−.519 3.162
11	2.714 16.48	858 4.018	.329 3.006	.050 2.866	−.128 2.920	−.254 3.027	−.348 3.146	−.422 3.261	−.481 3.371	−.530 3.470

The surfaces for μ_1', σ and α_4 are more irregular. The problem of determining a cumulative function with $\alpha_3 = a$ and $\alpha_4 = b$ is equivalent to the problem of determining a point of intersection of the curves

$$\alpha_3 = G(k, c), \qquad \alpha_3 = a \tag{58}$$
$$\alpha_4 = H(k, c), \qquad \alpha_4 = b.$$

Direct algebraic solution of this system appears very difficult, and other techniques must be resorted to.

One method is to use only integral values of k, and then for each k interpolate for the value of c giving the desired α_3. For such pairs of c and k, find α_4 by interpolation. Then choosing the pairs having α_4 just above and just below the desired one, the proper linear combination (51) is taken. This gives a combination function which has both α_3 and α_4 at the desired values. This combination

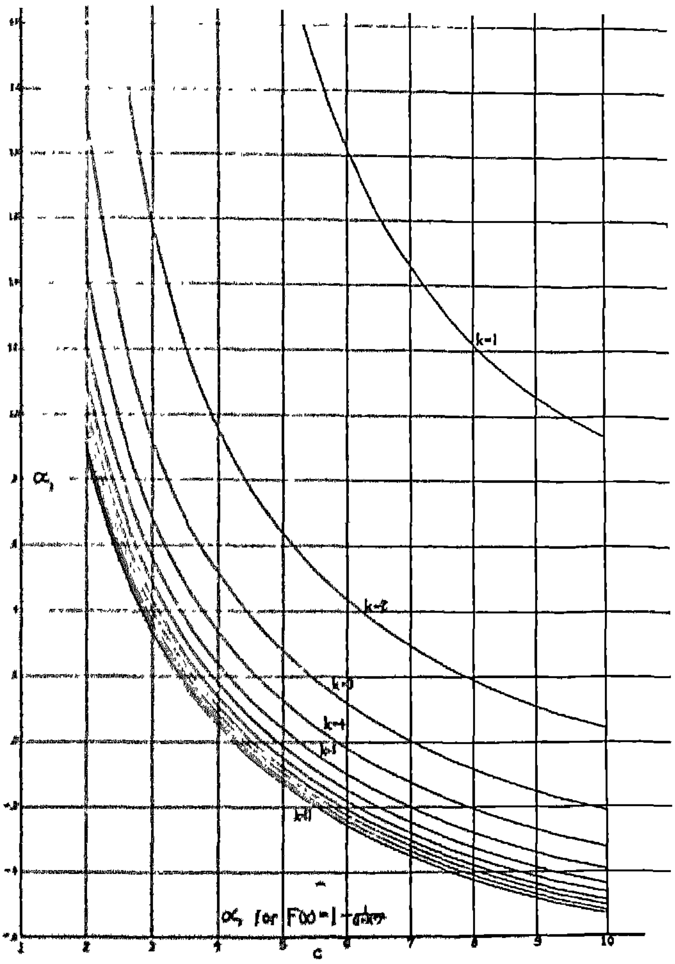


FIGURE I

will be an approximation to the single function with non-integral k, having the given α_3 *and* α_4. This method of linear combinations might be extended to fit α_5 by using three integral values of k.

The interpolations may be done graphically by use of Figure 1 and others like it. Or one may use Stirling's formula [5]. The interpolation for c from α_3 is backwards, while that for α_4 from c is direct. Sometimes it is more accurate

to use Newton's formula [5, p. 36] when the values in one direction increase rapidly.

Use of a single function $F(x)$ for a graduation is easily accomplished. First, obtain the c and k to be used so that α_3 is correct and α_4 is as close to the given value as possible. Then determine μ_1' and σ from Table II by interpolation. Change the scale and origin of the original values of the variable X to those x's corresponding to $F(x) = 1 - 1/(1 + x^c)^k$, through

$$\frac{x - \mu_1}{\sigma} = t = \frac{X - M}{S}, \tag{59}$$

where M and S are the mean and standard deviation of the given distribution. Now compute the values of $1/(1 + x^c)^k$ for the various values of x. The differences of these results are equal to the differences of $F(x)$, which by (1) are the probabilities for the given ranges of X. Multiplication by the total frequency will yield the theoretical frequencies, if desired.

If the graduation is to be done by a combination of two functions, the work is carried out for each as described above, and then the frequencies are combined by the same linear combination as that by which the component α_4's must be combined to give the desired α_4. This may readily be seen by considering the separate cumulative functions in terms of the standard variable t, whence the means and σ's are 0 and 1 and (51) is applicable. Then the differences of $G(t) = k_1G_1(t) + k_2G_2(t)$ are sought. But these can be found by taking the same linear combination of the separate differences of the functions $G_1(t)$ and $G_2(t)$. However, these values are merely computed from their respective sets of x values.

For illustration, three graduations are given. The first is a highly normal distribution of heights from Rietz [5, p. 98ff.]. For this distribution, $M = .02085$, $S = 2.5723$, $\alpha_3 = -.0124$, $\alpha_4 = 3.149$. The graduation was done by taking the function $F(x) = 1 - \dfrac{1}{(1 + x^6)^4}$ which has the nearly normal characteristics $\alpha_3 = -.019$, $\alpha_4 = 3.169$. The object was to take a simple cumulative function with integral k and c to show how a satisfactory job can be done on a normal distribution For this function $\mu_1' = .75550$ and $\sigma = .16234$. Then

$$x = .063110X + .75418,$$

into which are substituted the X class-limits -11.5, -10.5, etc. From these, corresponding values of $\dfrac{8585}{(1 + x^6)^4}$ are calculated and differenced to give the theoretical frequencies for the 8585 cases. The results are given in Table IV.

The fit obtained by use of $F(x)$ is good. One comparison test is that of χ^2. The eight classes -11, -10, -9, 9, 10, 11, 12, 13 were grouped together. The results were

$$\chi_F^2 = 21.210, \qquad \chi_V^2 = 23.479,$$

as compared to

$$P(\chi^2 > 22.31) = .10,$$
$$P(\chi^2 > 19.31) = .20,$$

for 15 degrees of freedom (18 classes minus 3 for linear restrictions). One reason for the somewhat lower χ^2 for $F(x)$ may be that its α_3 and α_4 are closer to

TABLE IV

X	Observed frequency [5]	Graduated frequency by $F(x)$	Graduated frequency of normal [5]
−11.0		00	16[3]
−10.0	2	.43	67
−9.0	4	3.23	2.84
−8.0	14	13.17	10.30
−7.0	41	39.81	32.11
−6.0	83	97.87	86.03
−5.0	169	206.72	198.17
−4.0	394	385.34	392.43
−3.0	669	639.55	668.11
−2.0	990	941.98	977.92
−1.0	1223	1216.47	1230.63
.0	1329	1353.98	1331.41
1.0	1230	1278.39	1238.41
2.0	1063	1013.80	990.33
3.0	946	676.12	680.86
4.0	392	384.41	402.44
5.0	202	190.83	204.51
6.0	79	85.19	89.35
7.0	32	35.24	33.56
8.0	16	13.87	10.84
9.0	5	5.33	3.01
10.0	2	2.01	72
11.0		.77	.15
12.0		.30	03
13.0		19[3]	
Total.	8585	8585.00	8584.99

those of the observed distribution than are those of the normal function. This gives a better fit in the tails of the distribution. Nevertheless, this example does illustrate how one of the simplest of the cumulative functions with "normal" characteristics can be used without specifically fitting α_3 and α_4. It may also be mentioned that $F(x)$ for $c = 5$, $k = 6$ has α_3 and α_4 even closer to the normal

[3] Total of stump frequency.

TABLE V

X	Observed frequency	$F(x)$, $k = 4$ $c = 3.228$, f_4	$F(x)$, $h = 5$ $c = 2.944$; f_5	$F(x) = .3063f_4 + .6937f_5$	Type III [5]
−8.0	3	.00	00	00	
−7.0	9	86	00	27	2
−6.0	46	39.58	25.07	29.52	27
−5.0	167	180.78	175.27	176.96	142
−4.0	372	433.86	445.79	442.13	410
−3.0	718	768.83	791.72	784.71	790
−2.0	1186	1116.06	1134.52	1128.86	1186
−1.0	1462	1383.06	1384.99	1384.40	1441
.0	1498	1492.04	1477.86	1482.20	1502
+1.0	1460	1419.70	1399.70	1405.83	1385
2.0	1142	1205.81	1190.80	1195.40	1158
3.0	913	926.59	921.47	923.01	891
4.0	642	654.00	656.82	655.96	641
5.0	435	430.66	436.78	434.90	434
6.0	235	268.70	274.63	272.81	280
7.0	167	161.10	165.27	163.99	173
8.0	133	93.99	96.23	95.55	102
9.0	47	53.88	54.77	54.50	59
10.0	29	30.62	30.70	30.68	33
11.0	13	17.37	17.07	17.16	18
12.0	9	9.86	9.46	9.58	9
13.0	5	5.64	5.26	5.38	5
14.0	8	3.26	2.93	3.03	2
15.0	2	1.89	1.66	1.73	1
16.0		1.12	.93	.99	1
17.0		.66	.53	57	
18.0		.41	.31	.34	
19.0		.24	.18	.20	
20.0		.16	.11	.13	
21.0		27[4]	.17[4]	.20[4]	
Total . . .	10701	10701.00	10701.00	10700.99	10701

TABLE VI

Observed [6]	Type III [6]	Type A [6]	Edgeworth [6]	$F(x)$
3	4	5	4	4
20	17	22	17	19
38	42	47	42	42
63	59	60	59	56
51	53	50	53	52
29	33	27	32	34
21	15	13	15	16
4	5	4	6	5
0	1	1	2	1
1	0	0	1	0
230	229	229	231	229
χ^2	4.54	7.55	5.86	4.03

[4] Stump frequency.

values, but it does not give quite as good a fit because it tends to decrease too rapidly on the left.

The second example is also from Rietz [5, p. 108ff.]. For this distribution, $M = .68835$, $S = 2.9180$, $\alpha_3 = .583$ and $\alpha_4 = 3.698$. Two functions were used with $k = 4$ and $k = 5$ By interpolation

		μ_1'	σ	α_3	α_4
$k = 5$	$c = 2.914$	.54200	.22247	.583	3.655
$k = 4$	$c = 3.228$	.61577	.23823	.583	3 795

Because of the rather rapid increases for smaller values of c, Newton's formula [5, p. 36] yields better approximations than Stirling's [5, p 38 (12)]. The graduation for each function is carried out as above, and since

$$.3063\cdot 3.795 + .6937\cdot 3.655 = 3.698,$$

the linear form

$$.3063f_4' + .6937f_5' = f',$$

is used.

Table V gives the component and combined frequencies, and also the frequencies from a Type III. χ^2 for both are very high even though the fit appears reasonably good on a graph. This result is due to classes 6 and 8 which tend to cause a high χ^2 for any distribution function of a small number of parameters. The example, however, does show that $F(x)$ can be used to graduate a skewed distribution

It is to be further noted that the component functions were used only to obtain an approximation to a single function with $4 < k < 5$, for which α_3 and α_4 are simultaneously correct. When tables more complete than Tables II and III are available, such a single function can be found

The third example of graduations is from Elderton [6]. The measures were treated as a discrete variable in computing α_3 and α_4. A single function $c = 3.102$, $k = 11$ was used. This function had α_3 at the observed value of .2936, while α_4 was 2.973 as compared to the observed 2.986. The results along with those by classical methods are shown in Table VI. The above χ^2 were obtained by grouping the first and the last three class frequencies. The values are approximate because of rounding. However, they do show that $F(x)$ does a comparable graduation.

Besides aiding in the problem of graduation, this cumulative function should prove of value in the approximation of known or population distributions, as for example, $(p + q)^n$. However much more work needs to be done before this can be more than a conjecture.

7. Conclusion. This paper has stressed the advantages obtained by the direct use of the cumulative function. A number of useful functions have been considered. A general method for fitting any cumulative function by the construction of a table has been suggested. A particular method depending

upon the use of certain new cumulative moments has been given. Making use of this theory a certain simple algebraic function has been discussed in detail, and its use in graduations explained.

The writer wishes to convey his sincere thanks to Professor Harry C. Carver, whose counsel was most helpful.

BIBLIOGRAPHY

[1] D. Bierens de Haan, *Nouvelles Tables d'Intégrales Définies*, 1867 (Corrected edition 1939), G. E. Stechert, New York.

[2] J. F. Steffensen, *Interpolation*, 1927, Williams & Wilkins Co., Baltimore, Md., pp. 57 ff.

[3] E. Gifford, *Natural Sines*, 1914, Abel Heywood, Manchester.

[4] J. Bauschinger, J. Peters, *Logarithmic-Trigonometric Tables with Eight Places*, I, II, 1911, Wilhelm Engelmann, Leipzig.

[5] H. L. Rietz, ed., *Handbook of Mathematical Statistics*, 1924, Houghton Mifflin, Cambridge, Mass.

[6] W. P. Elderton, *Frequency Curves and Correlation*, (3rd edition), 1938, Cambridge University Press, London, p. 134 ff.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

AN APPROXIMATE NORMALIZATION OF THE ANALYSIS OF VARIANCE DISTRIBUTION

BY EDWARD PAULSON[1]

Columbia University

The statistic $F = s_1^2/s_2^2$, where s_1^2 and s_2^2 are two independent estimates of the same variance, has played an essential part in modern statistical theory. All tests of significance involving the testing of a linear hypothesis, which includes the analysis of variance and covariance and multiple regression problems, can be reduced to finding the probability integral of the F distribution. This distribution (and the equivalent distribution of $z = \frac{1}{2}\log F$) has so far been directly tabulated only for the 20, 5, 1, and 0.1 percent levels of significance [1] To find the critical value of F for some other probability level would require the use of Pearson's extensive triple-entry tables [2], which is not very convenient to use for this purpose, and in addition is inadequate for some ranges of the parameters.

It therefore appears that it might be of some practical value to have an approximate method of determining the critical values of F for other probability levels. A solution will be given based on a modified statistic U, a function of F, so selected as to tend to have a nearly normal distribution with zero mean and unit variance. This normalized statistic will have the additional advantage that further tests are possible with normalized variates, as pointed out by Hotelling and Frankel [3].

F can be written in the form

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2},$$

where χ_1^2 and χ_2^2 have the chi-square distribution with n_1 and n_2 degrees of freedom respectively. It is known from the work of Wilson and Hilferty [4] that $\left(\frac{\chi^2}{n}\right)^{\frac{1}{3}}$ is nearly normally distributed with mean $1 - 2/9n$ and variance $2/9n$. An obvious approach to the problem of securing an approximation to the F distribution is to regard $F^{\frac{1}{3}}$ as the ratio of two normally distributed variates. In general the distribution of the ratio $r = y/x$ where y and x are normally and independently distributed with means m_y and m_x and standard deviations σ_y and σ_x

[1] Work done under a grant-in-aid from the Carnegie Corporation of New York

is not expressible in simple form. However Fieller [5] has shown that a function R of v, namely $R = \dfrac{vm_x - m_y}{\sqrt{v^2\sigma_x^2 + \sigma_y^2}}$ will be nearly normally distributed with zero mean and unit variance, provided the probability of x being negative is small. In the given problem it follows that we can regard

$$(1) \qquad U = \frac{\left(1 - \dfrac{2}{9n_2}\right)F^{\frac{1}{3}} - \left(1 - \dfrac{2}{9n_1}\right)}{\sqrt{\dfrac{2}{9n_2}F^{\frac{2}{3}} + \dfrac{2}{9n_1}}},$$

as nearly normally distributed (with zero mean and unit variance) provided $n_2 \geq 3$, for with $n_2 = 3$ the probability of the denominator of $F^{\frac{1}{3}}$ being negative is only .0003. If it is desired to use the lower tail of the F distribution, then the statistic U should only be used if n_1 is also ≥ 3. Ordinarily, in most applications only the upper tail of the F distribution is used, and n_2, which corresponds to the number of degrees of freedom in the estimate of the error variance, will be much greater than 3.

The following tables show the degree of accuracy of the approximation. The exact value of F corresponding to various levels of significance are compared

P	$n_1 = 1, \quad n_2 = 10$	
	$t = \sqrt{F}$	
	Approximation	Exact Value
.20	1.37	1.37
.05	2.21	2.23
.01	3.16	3.17
.001	4.63	4.59
.0001	6.40	6.22

P	$n_1 = 4, \quad n_2 = 8$		$n_1 = 6, \quad n_2 = 12$	
	F		F	
	Approximation	Exact Value	Approximation	Exact Value
.99	.058	.068	.123	.130
.95	.161	.166	.248	.250
.80	.407	.406	.497	.496
.20	1.92	1.92	1.72	1.72
.05	3.84	3.84	3.00	3.00
.01	7.12	7.01	4.85	4.82
.001	15.38	14.39	8.58	8.38

with the approximate values, which are found by solving (1) for F by considering it as a quadratic equation in $F^{\frac{1}{2}}$. In these tables $P = \int_{F}^{\infty} \varphi(F)\, dF$, where $\varphi(F)$ is the probability distribution of F. The case $n_1 = 1$ is of special interest, since here $F = t^2$, where t has Student's distribution, and is shown separately

REFERENCES

[1] R. A. FISHER and F YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*, London, 1938.

[2] KARL PEARSON (Editor), *Tables of the Incomplete Beta Function*, Biometric Laboratory, London, 1934

[3] H HOTELLING and L. R. FRANKEL, "The transformation of statistics to simplify their distribution," *Annals of Math Stat*, Vols 8–9 (1937–38), pp. 87–96.

[4] E. B WILSON and M. M. HILFERTY, "The distribution of chi-square," *National Acad Sc Proc*, Vol. 17 (1931), pp. 684–688

[5] E C FIELLER, "The distribution of the index in a normal bivariate population," *Biometrika*, Vol 24 (1932), pp 428–440.

NOTE ON THE DISTRIBUTION OF ROOTS OF A POLYNOMIAL WITH RANDOM COMPLEX COEFFICIENTS

BY M. A. GIRSHICK

United States Department of Agriculture

In order to obtain the distribution of roots of a polynomial with random complex coefficients, it was found convenient to employ a rather well known theorem on complex Jacobians. Since proofs of this theorem are not very plentiful in the literature, a brief and simple proof of it is presented in this note

THEOREM: *Let n analytic functions be defined by*

$$w_p = u_p + iv_p = f_p(z_1, z_2, \cdots, z_n), \quad (p = 1, 2, \cdots, n), \tag{1}$$

where $z_p = x_p + iy_p$, $i = \sqrt{-1}$. Let j denote the Jacobian of the transformation of the n complex variables defined by (1). *That is*

$$j = \begin{vmatrix} \frac{\partial w_1}{\partial z_1} & \cdots & \frac{\partial w_1}{\partial z_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial w_n}{\partial z_1} & \cdots & \frac{\partial w_n}{\partial z_n} \end{vmatrix}. \tag{2}$$

Let furthermore J denote the Jacobian of the transformation of the $2n$ real variables defined by the equations $u_p = u_p(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n)$ and $v_p = v_p(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n)$, $(p = 1, 2, \cdots, n)$. That is

$$J = \begin{vmatrix} U_x & U_y \\ V_x & V_y \end{vmatrix}, \tag{3}$$

where

$$U_x = \begin{Vmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial u_n}{\partial x_1} & \cdots & \frac{\partial u_n}{\partial x_n} \end{Vmatrix}, \qquad U_y = \begin{Vmatrix} \frac{\partial u_1}{\partial y_1} & \cdots & \frac{\partial u_1}{\partial y_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial u_n}{\partial y_1} & \cdots & \frac{\partial u_n}{\partial y_n} \end{Vmatrix},$$

$$V_x = \begin{Vmatrix} \frac{\partial v_1}{\partial x_1} & \cdots & \frac{\partial v_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial v_n}{\partial x_1} & \cdots & \frac{\partial v_n}{\partial x_n} \end{Vmatrix}, \qquad V_y = \begin{Vmatrix} \frac{\partial v_1}{\partial y_1} & \cdots & \frac{\partial v_1}{\partial y_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial v_n}{\partial y_1} & \cdots & \frac{\partial v_n}{\partial y_n} \end{Vmatrix}.$$

Then J equals the square of the modulus of j.

PROOF: Since by hypothesis w_p is analytic we can set $\frac{\partial w_p}{\partial z_q} = \frac{\partial v_p}{\partial y_q} - i\frac{\partial u_p}{\partial y_q}$. Hence j takes on the form:

$$j = | V_y - iU_y |. \tag{4}$$

Again, since w_p is analytic, we have $\frac{\partial u_p}{\partial x_q} = \frac{\partial v_p}{\partial y_q}, \frac{\partial v_p}{\partial x_q} = -\frac{\partial u_p}{\partial y_q}$. That is $U_x = V_y$ and $V_x = -U_y$. Hence J in (3) has the value

$$J = \begin{vmatrix} V_y & U_y \\ -U_y & V_y \end{vmatrix}. \tag{5}$$

Now J can also be written in the form

$$J = \begin{vmatrix} V_y & iU_y \\ iU_y & V_y \end{vmatrix}. \tag{6}$$

This follows from the fact that if we multiply each of the last n rows of the expression for J in (6) by i and factor out i from the last n columns, we get the expression for J given in (5).

Now in (6) subtract the $(n + p)$th row from the pth row for each $p = 1, 2, \cdots, n$. This yields:

$$J = \begin{vmatrix} V_y - iU_y & iU_y - V_y \\ iU_y & V_y \end{vmatrix}. \tag{7}$$

Next add in (7) the pth column to the $(n + p)$th column for each $p = 1, 2, \cdots, n$ This yields:

$$J = \begin{vmatrix} V - iU & 0 \\ iU & V + iU \end{vmatrix} = | V - iU || V + iU |. \tag{8}$$

But (8) is precisely the square of the modulus of $| V - iU |$. This in conjunction with (4) proves the theorem.

Consider the equation

$$z^n - a_1 z^{n-1} + \cdots + (-1)^n a_n = 0, \tag{9}$$

where the a_i are complex numbers. We may wish to consider the real and imaginary parts of a_i as random variables having a given joint distribution function, and require to find the probability that one or more roots of (9) will lie in a specified region of the complex plane. In order to answer this question, it is necessary to find the joint distribution of the real and imaginary parts of the roots of (9).

As an example let us assume that the real and imaginary parts of a_p are normally and independently distributed with zero mean and variance σ^2. That is, we assume that the distribution density of these quantities is given by

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{2n} \exp\left[-\frac{1}{2\sigma^2}\sum_{p=1}^{n} a_p \bar{a}_p\right] \tag{10}$$

where $\bar{a}_p$ is the conjugate of a_p. Let $z_1, z_2, \cdots, z_n$ be the roots of (9). The relationship between the roots and coefficients of (9) are given by

$$a_1 = \sum_{j=1}^{n} z_j, \qquad a_2 = \sum_{j<k} z_j z_k, \cdots, \qquad a_n = z_1 z_2 \cdots z_n \tag{11}$$

Thus the a_p's are analytic functions of the z's.

In order to find the joint distribution of the real and imaginary parts of the z's, it is necessary to find the real Jacobian J of the transformation defined by (11). Now the complex Jacobian j of the transformation (11) is defined as

$$j = \begin{vmatrix} \frac{\partial a_1}{\partial z_1} & \cdots & \frac{\partial a_1}{\partial z_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial a_n}{\partial z_1} & \cdots & \frac{\partial a_n}{\partial z_n} \end{vmatrix}. \tag{12}$$

A simple calculation will show that the value of j in (12) is given by

$$j = \sum_{p=1}^{n} \sum_{q=p+1}^{n} (z_p - z_q). \tag{13}$$

Hence, applying the theorem proved above, we get

$$J = | j |^2 = \sum_{p=1}^{n} \sum_{q=p+1}^{n} | z_p - z_q |^2, \tag{14}$$

where the symbol $||$ stands for the modulus.

From (10) and (14) we conclude that the joint distribution density of the real and imaginary parts of the roots of (9) is given by

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{2n} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{n} z_j \sum_{j=1}^{n} \bar{z}_j + \cdots + z_1\bar{z}_1 \cdots z_n\bar{z}_n\right\}\right] \sum_{p=1}^{n} \sum_{q=p+1}^{n} |z_p - z_q|^2. \tag{15}$$

A NOTE ON THE PROBABILITY OF ARBITRARY EVENTS

By Hilda Geiringer[1]

Bryn Mawr College

In a recently published paper [1] on arbitrary events the author studies the probability of the occurrence of at least m among n events. Denoting by $p_m(\gamma_1, \gamma_2, \cdots \gamma_r)$ the probability that at least m among the r events, $E_{\gamma_1}, \cdots E_{\gamma_r}$ occur, and by $p_{(\alpha_1, \alpha_2, \ldots, \alpha_r)}$ the probability of the non occurrence of the events numbered $\alpha_1, \alpha_2, \cdots \alpha_r$ and of the occurrence of the $n - r$ others, he proves

$$\begin{aligned} -p_1(\alpha_{r+1}, \cdots \alpha_n) + \sum_{\gamma_1} p_1(\gamma_1, \alpha_{r+1}, \cdots \alpha_n) - \sum_{\gamma_1} \sum_{\gamma_2} p_1(\gamma_1, \gamma_2, \alpha_{r+1}, \cdots \alpha_n) \\ + \cdots + (-1)^r \sum p_1(1, \cdots n) = p_{(\alpha_1, \cdots \alpha_r)} . \end{aligned} \tag{I}$$

(Theorem VI, page 336). From (I) he deduces that *a necessary and sufficient condition* for the existence of a system of events $E_1, \cdots E_n$ associated with given values $t_1(\alpha_1, \cdots \alpha_k)$ is that the expressions on the left side of (I) computed from these t's are $\geqq 0$ for all possible combinations of the α's (Theorem VII). He also points out that it was not possible to find similar (necessary and sufficient) conditions for $m \neq 1$. I wish to show in this note the relation between these theorems and some well known basic facts of the theory of arbitrarily linked events and to add some remarks.

1. Given n chance variables x_i $(i = 1, \cdots n)$ denote by $x_i = 1$ the "occurrence of E_i", by $x_i = 0$ its non occurrence and by $v(x_1, x_2, \cdots x_n)$ the probability of "the result $(x_1, x_2, \cdots x_n)$" i.e., the probability that the first variable equals x_1 the second x_2, $\cdots$ the last x_n; e.g. $v(1, 1, 1, 0, \cdots 0) = v_{[45 \ldots n]}$ is the probability that only the three first events occur. Hence the v's are 2^n probabilities, arbitrary except for the *condition to have the sum* 1.

Instead of these v's we often introduce another set of $2^n - 1$ probabilities, namely p_i the probability of the occurrence of E_i $(i = 1, \cdots n)$; p_{ij} that of the joint occurrence of E_i and E_j $(i, j = 1, \cdots n)$; $\cdots$ $p_{12 \ldots n}$ the probability that *all* the events occur.

---

[1] Research under a grant-in-aid of the American Philosophical Society.

It may be noted that instead of the $p_i, p_{ij}, \cdots p_{12\ldots n}$ we could quite as well use a system of $q_i, q_{ij}, \cdots q_{12\cdots n}$ where q_i is the probability of the non-occurrence of E_i (or of the occurrence of $E_i' = E - E_i$, q_{ij} that of the joint non-occurrence of E_i and E_j (of the occurrence of $E_i'E_j'$) and $q_{12\cdots n}$ the probability of $E_1'E_2' \cdots E_n'$.

The use of the p's (or q's) instead of the "elementary probabilities" v is justified by the fact that the p's are $(2^n - 1)$ *independent* linear combinations of the v's and that therefore *the v's and the p's* (or the v's and the q's) *determine each other uniquely*. There exist in fact the following well known relations, (1) and (2). The first set (1) gives just the definition of the $2^n - 1$ probabilities p, in terms of the v's, and the second set expresses the v's by the p's as the result of the solution of the $2^n - 1$ independent linear equations (1). Thus we have, beginning with $p_{12\ldots n}$.

$$
(1)\qquad
\begin{aligned}
p_{12\cdots n} &= v(1, 1, \cdots 1),\\
p_{12\cdots n-1} &= \sum_{x_n} v(1, 1, \cdots 1, x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
p_{12} &= \sum_{x_3} \cdots \sum_{x_n} v(1, 1, x_3, x_4, \cdots x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
p_n &= \sum_{x_1} \cdots \sum_{x_{n-1}} v(x_1, x_2, \cdots x_{n-1}, 1),
\end{aligned}
$$

and solving successively:

$$
(2)\qquad
\begin{aligned}
v(1, 1, \cdots 1) &= p_{12\cdots n}\\
v(1, 1, \cdots 1, 0) &= p_{12\cdots n-1} - p_{12\cdots n},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
v(1, 1, 0, \cdots 0) &= p_{12} - \sum_{\gamma_1} p_{12\gamma_1} + \sum_{\gamma_1}\sum_{\gamma_2} p_{12\gamma_1\gamma_2} - \cdots \pm p_{12\cdots n},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
v(0, 0, \cdots 0, 1) &= p_n - \sum_{\gamma_1} p_{\gamma_1 n} + \sum_{\gamma_1}\sum_{\gamma_2} p_{\gamma_1\gamma_2 n} - \cdots\\
&\qquad \pm \sum_{\gamma_1} \cdots \sum_{\gamma_{n-1}} p_{\gamma_1\gamma_2\cdots\gamma_{n-1}n} \mp p_{12\cdots n}.
\end{aligned}
$$

The successive solution of the system (1) with respect to the "unknowns" v is possible because each new equation in (1) contains *exactly one* new unknown v; e.g. in the equation defining p_{12} the only "unknown" is $v(1, 1, 0, 0, \cdots 0)$ all the v's with more than two "1"s having already been computed from the foregoing equations.

If we choose to use the system of the q's we have in the same way:

$$
(1')\qquad
\begin{aligned}
q_{12\cdots n} &= v(0, 0, \cdots 0),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
q_n &= \sum_{x_1} \cdots \sum_{x_{n-1}} v(x_1, x_2, \cdots x_{n-1}, 0),
\end{aligned}
$$

and the inverse system

$$(2')\qquad \begin{aligned} v(0, 0, \cdots 0) &= q_{12\cdots n} \\ v(0, 0, \cdots 0, 1) &= q_{12\cdots n-1} - q_{12\cdots n} \\ &\cdots\cdots\cdots\cdots \\ v(1, 1, \cdots 1, 0) &= q_n - \sum_{\gamma_1} q_{\gamma_1 n} + \sum_{\gamma_1}\sum_{\gamma_2} q_{\gamma_1\gamma_2 n} - \cdots \mp q_{12\cdots n}. \end{aligned}$$

Coming back to Chung's theorem we see that the probability $p_1(\alpha_1, \cdots \alpha_r)$ that at least one event among $E_{\alpha_1}, \cdots E_{\alpha_r}$ occurs is evidently:

$$(3)\qquad p_1(\alpha_1, \cdots \alpha_r) = 1 - q_{\alpha_1,\ldots\alpha_r}.$$

If we introduce this value in (I) all the "1"s introduced by (3) cancel and *we get our system* (2′). (Of course we could in the same way deduce from (2) a system of equations for $q_1(\alpha_1, \cdots \alpha_r) = 1 - p_{\alpha_1\ldots\alpha_r}$ where $q_1(\alpha_1, \cdots \alpha_r$ is the probability that at most $r - 1$ events among the r given ones occur.)

As the v-values on the left side of (2′) are $2^n - 1$ independent probabilities only subject to the restriction that they have the sum $\leqq 1$, we see that the expressions on the right side of (2′) *must have the same properties*; and these properties are also *sufficient* for a system of q's, (or for a system of $p_1(\alpha_1, \cdots \alpha_r)$); indeed *if* they are fulfilled, these $2^n - 1$ expressions define by means of (2′) a system of elementary probabilities $v(x_1, \cdots x_n)$. Hence the theorems VI and VII quoted at the beginning of this note are rather close consequences of the basic relations of the theory of arbitrary events.

2. Remark 1. We may add one more equation to equations (1), namely

$$1 = \sum_{x_1} \cdots \sum_{x_n} v(x_1, \cdots x_n),$$

thus introducing $v(0, 0, \cdots 0)$. Then in system (2) the corresponding new equation will be:

$$v(0, 0, \cdots 0) = 1 - \sum_{\gamma_1} p_{\gamma_1} + \sum_{\gamma_1}\sum_{\gamma_2} p_{\gamma_1\gamma_2} - \cdots \pm p_{12\cdots n}$$

(and analogously for the q's). In this way we get two systems $(\bar{1})$ and $(\bar{2})$ each consisting of 2^n equations and in $(\bar{2})$ the sum of the expressions on the right side is now *identically equal* to one. Hence necessary and sufficient conditions will now be that *all these* 2^n *expressions must be non-negative.*

Remark 2. It is convenient to interpret or prove results of the kind considered here in terms of elementary measure theory: p_1 is the measure of a set E_1; p_2 of E_2; p_{12} that of the intersection E_1E_2 etc., and analogously for the v's: e.g. $v(1, 1, 0, \cdots 0) = m(E_1E_2E_3' \cdots E_n')$. Consider now the equations (2). The first is an identity. In the second $p_{12\ldots n-1}$ measures the product of $E_1E_2 \cdots E_{n-1}$, whereas $p_{12\cdots n-1} - p_{12\ldots n}$ is the measure of *that* part of this product which does not belong to E_n, and it therefore equals $m(E_1E_2 \cdots E_{n-1}E_n') = v(1, 1 \cdots 1, 0)$. In the last equation (2) $\sum_{\gamma_1} \cdots \sum_{\gamma_{n-2}} p_{\gamma_1\cdots\gamma_{n-2}n}$ is the

measure of that part of E_n which belongs to at least $(n - 2)$ other sets (besides E_n); whereas this same value minus $p_{12..n}$ is the measure of the part of E_n which belongs *exactly* to $(n - 2)$ other sets; now subtracting this expression from $\sum_{\gamma_1} \cdots \sum_{\gamma_{n-3}} p_{\gamma_1 \cdots \gamma_{n-3}}$ we get the measure of the part of E_n which belongs *exactly* to $(n - 3)$ other sets and finally $p_n - \cdots \mp p_{12..n}$ is the measure of that part of E_n which belongs to *no* other set besides, i.e $m(E_1'E_2' \cdots E_{n-1}'E_n) = v(0, 0, \cdots 0, 1)$. This kind of proof does not require the solution of (1).

REMARK 3. According to (1) the $p_i, p_{ij}, \cdots p_{12..n}$ are the ordinary *moments* of order $1, 2 \cdots n$ of $v(x_1, x_2, \cdots x_n)$. There are of course many more than $2^n - 1$ moments of this n-variate distribution but only $2^n - 1$ of them are different from each other because $1^r = 1$

3. Denote by $p_n(x)$, $(x = 0, 1, \cdots n)$ the *probability of getting exactly x successes in n trials.* (See e g. [2], [3].) For the simplest case of arbitrary events, the Bernoulli problem, $p_n(x) = \binom{n}{x} p^x(1 - p)^{n-x}$. Then the probabiliy of *at least* x successes (of a number of successes $\geqq x$) is

$$V_n(x) = p_n(x) + p_n(x + 1) + \cdots p_n(n), \tag{4}$$

or $p_x(1, 2, \cdots n)$ in Chung's notation The $p_n(x)$ are by their *definition* $(n + 1)$ *arbitrary positive* numbers with *sum equal to one.* These are the only *necessary* and *sufficient* restrictions for $p_n(x)$. $V_n(x)$ the "cumulative" distribution of $p_n(x)$ which is defined for x between $(-\infty$ and $+\infty)$ is a monotone non-increasing *step function* with its $(n + 1)$ steps at $x = 0, 1, 2, \cdots n$ equal to the $p_n(x)$.

Consider next $p_x(\alpha_1, \alpha_2, \cdots \alpha_r)$ where $r < n$; these are cumulative distributions each corresponding to one of the $\binom{n}{r}$ probabilities $p_r(x)$ where $p_r(x)$ is the probability of exactly x successes in a group of r trials.[2] For each group $(\alpha_1, \cdots \alpha_r)$ the corresponding $p_r(x)$, $(x = 0, 1, \cdots r)$ are positive and with sum equal to one. Hence if we always omit $p_r(0)$ because of $\sum_x^{0 \; r} p_r(x) = 1$, all the different $p_1(x), p_2(x), \cdots p_n(x)$ together define

$$1 \cdot n + n(n - 1) + \binom{n}{2}(n - 2) + \cdots + n = n2^{n-1}$$

values. As $n2^{n-1} > 2^n$ for $n > 2$ we realise that between these $n2^{n-1}$ probabilities there must exist a set of $n2^{n-1} - (2^n - 1)$ *identical relations*; and the same is true for the corresponding cumulative distributions $V_r(x)$ or $p_x(\alpha_1, \cdots \alpha_r)$. Thus it seems reasonable that it may be hard to use these $p_x(\alpha_1, \cdots \alpha_r)$ in the characterization of a problem of arbitrarily linked events if $x > 1$. On the other hand we have seen in **1** that for $x = 1$ they reduce to the

[2] One may write here $p_r(x)$ instead of $p_{(\alpha_1, \alpha_2, \cdots, \alpha_r)}(x)$

$2^n - 1$ probabilities $q_i, q_{ij}, \cdots q_{12\cdots n}$ which of course define the system of events unequivocally.

4. Introduce in the usual way the *sums* of the p_i, p_{ij}, etc.

(5) $$S_1 = \sum_i p_i, \qquad S_2 = \sum_{i,j} p_{ij}, \cdots S_n = p_{12\cdots n}, \quad \text{and} \quad S_0 = 1.$$

Now add in system (1) first the n equations which define $p_1, p_2, \cdots p_n$, then the $\binom{n}{2}$ equations for the p_{ij}, etc Observing that $p_n(x)$ is the sum of all *these* elementary probabilities $v(x_1 x_2 \cdots x_n)$ with exactly x "1"s and $(n - x)$ "0"s we get as the result of these n additions the well known formulae:

(6) $$S_\gamma = \sum_{x=\gamma}^{n} \binom{x}{\gamma} p_n(x), \qquad (\gamma = 0, 1, \cdots n).$$

Here $\gamma = 0$ gives $S_0 = 1 = \sum_0^n p_n(x)$. We may solve successively these $(n + 1)$ linear equations with respect to $p_n(n), p_n(n - 1), \cdots p_n(0)$, each linear equation containing only one new unknown, and find:

(7) $$p_n(x) = \sum_{\gamma=x}^{n} (-1)^{\gamma+x} \binom{\gamma}{x} S_\gamma, \qquad (x = 0, 1, \cdots n).$$

(These formulae could also have been derived from (2) by collecting groups of equations such that all the corresponding $v(x_1, \cdots x_n)$ contain the same number of "1"s.) (In the measure interpretation $p_n(x)$ is the measure of that part which belongs exactly to x of the original sets and S_γ measures the set which belongs to at least γ of these sets.) We also find by "cumulating" equations (9)

(8) $$S_\gamma = \sum_{x=\gamma}^{n} \binom{x-1}{\gamma-1} V_n(x), \qquad (\gamma = 1, 2, \cdots n),$$

and the inverse system

(9) $$V_n(x) = \sum_{\gamma=x}^{n} (-1)^{\gamma+x} \binom{\gamma-1}{x-1} S_\gamma, \qquad (x = 1, 2, \cdots n).$$

(6) and (8) are of the same type as (1), and (7) and (9) of the same as (2). We also may deduce analogous formulae by interchanging the roles of 0 and 1 and introducing a system of $T_1, T_2, \cdots T_n$ which depends on the q's in the same way as the $S_1, S_2, \cdots S_n$ defined in (5) depend on the p's.

We have seen that the $p_n(x)$ are $(n + 1)$ arbitrary non-negative numbers subject to only the condition of having their sum equal to one. But the S_γ $(\gamma = 0, \cdots n)$ are not arbitrary as we see from (7). The $(n + 1)$ expressions on the right side of (7) *must each be non-negative if* they are to define the probabilities $p_n(x)$ (their sum is identically equal to one). *Then and only then* they define a system of arbitrarily liked events $E_1, \cdots E_n$.

The $p_n(x)$, $(x = 0, 1, \cdots n)$ are of course not equivalent to the complete

system of $2^n - 1$ values $v(x_1, x_2, \cdots x_n)$ and the same remark holds for the $S_0, \cdots S_n$ and the system of $p_i, p_{ij}, \cdots p_{12 \cdots n}$. But often we are particularly interested in problems dealing only with the $p_n(x)$ (and S_γ). (For instance the author has studied [2] the asymptotic behavior of $p_n(x)$ as n tends in different ways towards infinity.) The simplest way to indicate a particular p-system corresponding to given S_γ is of course to assume all the p_i equal to each other, all the p_{ij} equal to each other etc. and to put therefore:

$$p_1 = p_2 \cdots = p_n = \frac{1}{n} S_1,$$

$$p_{12} = \cdots = p_{n-1,n} = \left[1 \Big/ \binom{n}{2} \right] S_2, \cdots p_{12 \cdots n} = S_n .$$

In the corresponding v-system all these v's which show the same number of "1"s equal each other.

We see from (6) that the S_γ (multiplied by $\gamma!$) are the *factorial moments* of order $0, 1, \cdots n$ *of the distribution* $p_n(x)$. Therefore by (7) we get the $p_n(x)$ *in terms of their factorial moments up to order* n. We may therefore also say: *Necessary and sufficient conditions that a system of numbers* $N_0 = 1, N_1, \cdots N_n$ *be the factorial moment of an arithmetical distribution with at most* $(n + 1)$ *steps at* $x = 0, 1, \cdots n$ *are the inequalities.*

$$(10) \qquad \sum_{\gamma}^{x \cdots n} \frac{(-1)^{\gamma+x}}{x!\,(\gamma - x)!} N_\gamma \geqq 0, \qquad (x = 0, 1, \cdots n).$$

Note that here there is no more allusion to a set of arbitrary events; (10) are the necessary and sufficient conditions for a set of $(n + 1)$ numbers to be the $(n + 1)$ (factorial) moments of an arbitrary arithmetic distribution with its abscissae given. The linear inequalities (10) differ very much from the basic inequalities in the classical problem of moments; because in our problem the abscissae of the steps are given in advance.

5. In some problems (e.g., some questions connected with the law of large numbers, with correlation theory, with analysis of variance) we are only concerned with the *first and second* moment of a distribution. Thus we are lead to the following question: Given $r + 1$ numbers $N_0, N_1, \cdots N_r, (r \leqq n)$ indicate a set of necessary and sufficient conditions such that these numbers are the moments of an arithmetic distribution with at most $(n + 1)$ steps, at $0, 1, 2, \cdots n$.[3] Some sort of an answer which may work well in particular cases, can immediately be deduced from (10). "$r + 1$ numbers $N_0, N_1, \cdots N_r$ will be the factorial moments of an arithmetic distribution with, at most, $(n + 1)$ steps at $0, 1, 2, \cdots n$ if and only if it is possible to indicate s

[3] This problem and the method of its solution has much in common with a problem studied in R. von Mises' paper [4].

numbers $N_{r+1}, \cdots N_{r+s}$, $(0 \leqq s \leqq n - r)$, such that for the $r + s + 1$ numbers $N_0, N_1, \cdots N_r, \cdots N_{r+s}$ the $r + s + 1$ inequalities

$$\sum_{\gamma=x}^{r+s} \frac{(-1)^{\gamma+x}}{x!\,(\gamma - x)!} N_\gamma \geqq 0 \qquad (x = 0, 1, \cdots r + s)$$

be satisfied."

The proof of this statement is self evident but the statement itself cannot be considered satisfactory We get a general solution in the following way.

Let $f_1(t), \cdots f_r(t)$ be r functions of the chance variable t, $v(t)$ an arithmetic probability with n given attributes $t_1, t_2, \cdots t_n$ and

$$(11) \qquad E[f_\rho(t)] = \sum_1^n f_\rho(t_\gamma)v(t_\gamma) \equiv \sum_{\gamma=1}^n a_{\gamma\rho}v_\gamma = S_\rho, \qquad (\rho = 1, 2, \cdots r),$$

the expectations of $f_\rho(t)$ with respect to $v(t)$. We wish to indicate necessary and sufficient conditions for the r numbers S_ρ. For $f_\rho(t) = t^\rho$ we have the problem stated above where the first r moments are given.

Call (S) the r-dimensional curve $x_\rho = f_\rho(t)$ and $P_1, P_2, \cdots P_n$ the points on (S) with coordinates $f_\rho(t_\gamma) = a_{\gamma\rho}$, $(\rho = 1, \cdots r; \gamma = 1, \cdots n)$, S the given point with coordinates S_ρ. In this case, *the point S must be contained in the smallest convex body (B) determined by the n points $P_1, \cdots P_n$. This condition is necessary and sufficient.* Because, if we interpret the v_γ which are $\geqq 0$ as masses of the points P_γ, with sum equal to one, then S is the center of gravity of these masses and it is well known that the above mentioned condition for S has to be fulfilled. But this condition is also sufficient, because if S is contained in (B) there exists always a simplex of at most r dimensions, consisting of at most $(r + 1)$ of the given points such that S is the center of gravity of appropriate masses in these points.

If we want to indicate explicitly the inequalities for the S_ρ we must know the boundary of (B). This is determined by its *planes of support* ("Stützebenen," Minkowski) sometimes called *tack planes*. A tack plane is a plane which *does not separate* any two points of the given point set and contains at least one point of this set. A plane is said to *separate* two points if, when the coordinates of the points are written in the equation of the plane two values with opposite signs result. These definitions enable us to find those points P_γ which lie on the boundary of (B) and to determine this boundary. (E.g. for $r = 3$ we have to find such triples of i, k, l, that the determinant which represents the equation of the plane through these three points has the same sign for all possible other points P_λ. If the S_ρ are the first three *moments* with respect to the origin, these determinants become Vandermond determinants and we find easily that the boundary planes are each passing through two neighboring points P_γ, $P_{\gamma+1}$ and one of the endpoints P_1 or P_n. If $\rho = 2$, and the first two moments are given, the boundary of (B) consists of the polygon $P_1P_2 \cdots P_nP_1$) Then we find without difficulty the conditions to be satisfied by S in the form of *linear inequalities* between the given $S_1, S_2, \cdots S_r$.

We get the continuous case as a limit of the discontinuous case as $t_\gamma \to t_{\gamma+1}$

and the points $P(t)$ take up the whole curve (C), e g between $t = 0$ and ∞. Then the relations between the given S_ρ become *non-linear* inequalities, well known for the problem of moments.

REFERENCES

[1] KAI LAI CHUNG, "On the probability of the occurrence of at least m events among n arbitrary events," *Annals of Math. Stat.*, Vol 12 (1941).

[2] HILDA GEIRINGER, "Sur les variables aléatoires arbitrairement liees," *Revue Interbalcanique*, 1938.

[3] HILDA GEIRINGER, "Bemerkung zur Wahrscheinlichkeit nicht unabhangiger Ereignisse," *Revue Interbalcanique*, 1939

[4] R. v MISES, "The limits of a distribution function if two expected values are given," *Annals of Math Stat*, Vol 10, 1939

AN INEQUALITY FOR MILL'S RATIO

By Z. W. BIRNBAUM

University of Washington

Mr. R. D. Gordon[1] recently proved the inequalities

$$\frac{x}{x^2+1}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \leq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt \leq \frac{1}{x}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \quad \text{for } x > 0.$$

In the present note we show that the lower inequality can be replaced by the better estimate

$$\frac{\sqrt{4+x^2}-x}{2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \leq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt.$$

PROOF: According to a well-known theorem of Jensen[2], for $f(t)$ convex and $g(t) \geqq 0$ in the interval (a, b), the following inequality holds

$$f\left[\int_a^b tg(t)\,dt\Big/\int_a^b g(t)\,dt\right] \leq \int_a^b f(t)\,g(t)\,dt\Big/\int_a^b g(t)\,dt.$$

For $a = x, b = \infty, f(t) = 1/t, g(t) = te^{-t^2/2}$, this inequality gives

$$\int_x^\infty te^{-\frac{1}{2}t^2}\,dt\Big/\int_x^\infty t^2e^{-\frac{1}{2}t^2}\,dt \leq \int_x^\infty e^{-\frac{1}{2}t^2}\,dt\Big/\int_x^\infty te^{-\frac{1}{2}t^2}\,dt.$$

Since

$$\int_x^\infty te^{-\frac{1}{2}t^2}\,dt = e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \int_x^\infty t^2e^{-\frac{1}{2}t^2}\,dt = xe^{-\frac{1}{2}x^2} + \int_x^\infty e^{-\frac{1}{2}t^2}\,dt,$$

[1] R D. Gordon, "Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument," *Annals of Math. Stat*, Vol. 12 (1941), pp 364–366.

[2] See for example. G. H. Hardy, J E Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934, p 150–151.

we find

$$(e^{-\frac{1}{2}x^2})^2 \leq xe^{-\frac{1}{2}x^2} \int^{\infty} e^{-\frac{1}{2}t^2}\,dt + \left(\int_x^{\infty} e^{-\frac{1}{2}t^2}\,dt\right)^2,$$

and hence

$$\frac{\sqrt{4+x^2}-x}{2} \cdot e^{-\frac{1}{2}x^2} \leq \int_x^{\infty} e^{-\frac{1}{2}t^2}\,dt.$$

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIII, No. 3 — September, 1942

THE ANNALS
OF MATHEMATICAL STATISTICS

THE ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

ADDITIVE PARTITION FUNCTIONS AND A CLASS OF STATISTICAL HYPOTHESES

By J. Wolfowitz

New York City

1. Introduction. The purpose of the first part of this paper is to prove several theorems about a class of functions of partitions which are additive in structure and subject to mild restrictions These theorems may be regarded as contributions to the theory of numbers, but if one makes certain assignments of probabilities to the partitions the theorems may be expressed as statements about asymptotic distributions It is in this latter, probabilistic language, that we shall carry out the proofs, for the following reasons. The discussion will be more concise and certain circumlocutions will be avoided The theorems have statistical application and a number of theorems discussed recently in statistical literature are corollaries of one of our theorems.

In the second part of this paper the theory of testing statistical hypotheses where the form of the distribution functions is totally unknown and only continuity is assumed, will be discussed. The exact extension of the likelihood ratio criterion to this case will be given. Approximations to the application of this criterion in two problems will be proposed, one of which applies the results mentioned above. Lastly, in connection with the second problem, a combinatorial problem will be solved which is new and has interest per se.

2. Partitions of a single integer. Let n be a positive integer and $A = (a_1, a_2, \cdots, a_s)$ be any sequence of positive integers a_i $(i = 1, 2, \cdots, s)$, where $\sum_{i=1}^{s} a_i = n$, and s may be any integer from 1 to n. Two sequences A which have different elements or the same elements arranged in different order are to be considered distinct, so it is easy to see that there are 2^{n-1} sequences A. We shall consider the sequence A as a stochastic variable and assign to all sequences A the same probability, which is therefore 2^{-n+1}. Let r_j be the number of elements a in A which equal j $(j = 1, 2, \cdots, n)$, so that r_j is a stochastic variable. Let k be an integer $\leq n$. Then the joint distribution of the stochastic variables $r_1, r_2, \cdots, r_k$ is given as follows: The probability that $r_i = b_i$ $(i = 1, 2, \cdots, k)$ is

$$2^{-n+1}\left(\sum_{r=1}^{n} \sum \frac{r!}{(b_1)!(b_2)!\cdots(b_k)!(r_{(k+1)})!\cdots(r_n)!}\right), \tag{2.1}$$

where the inner summation is carried out over all sets of non-negative integers $r_{(k+1)}, \cdots, r_n$ such that

$$b_1 + b_2 + \cdots + b_k + r_{(k+1)} + \cdots + r_n = r, \tag{2.2}$$

$$b_1 + 2b_2 + \cdots + kb_k + (k+1)r_{(k+1)} + \cdots + nr_n = n \tag{2.3}$$

(The b_i, of course, are non-negative integers.)

Let $r = \sum_{i=1}^{n} r_i$, and

$$r'_{(k+1)} = \sum_{i=k+1}^{n} r_i \qquad (k < n),$$

so that r and r'_{k+1} are both stochastic variables. The probability that at the same time

$$r_i = b_i, \qquad (i = 1, \cdots, k), \tag{2.4}$$

and

$$r'_{(k+1)} = b'_{(k+1)}, \tag{2.5}$$

is given by (2.1) with the restriction

$$r_{(k+1)} + \cdots + r_n = b'_{(k+1)}, \tag{2.6}$$

added to the restrictions (2.2) and (2.3). With this added restriction the summation in (2.1) may be performed as follows: Let $t = \sum_{i=1}^{k} ib_i$. It is easy to see that the number of sequences A where every $a_i > k$, $r = r'_{(k+1)} = b'_{(k+1)}$, and $\Sigma a_i = n - t$, is given by the coefficient of x^{n-t} in the purely formal expansion in x of

$$(x^{k+1} + x^{k+2} + x^{k+3} + \cdots)^{b'_{(k+1)}} = x^{(k+1)b'_{(k+1)}}\left(\frac{1}{1-x}\right)^{b'_{k+1}},$$

and is

$$\binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}.$$

Hence $P\{(2.4) \text{ and } (2.5)\}$, where this symbol will always denote the probability of the relation in braces, is seen to be

$$2^{-n+1}\frac{\left(\sum_{i=1}^{k} b_i + b'_{(k+1)}\right)!}{(b'_{(k+1)})!\prod_{i=1}^{k}(b_i)!}\binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}. \tag{2.7}$$

If X is a stochastic variable, let $E(X)$ and $\sigma^2(X)$ denote, respectively, the mean and variance of X (if they exist), and if Y is another stochastic variable, let $\sigma(XY)$ be the covariance between X and Y. Also let $\bar{X} = \dfrac{X - E(X)}{\sigma(X)}$. By the distribution of X we shall mean a function $\varphi(x)$ such that $P\{X < x\} = \varphi(x)$. These conventions being established, we seek first to evaluate $E(r_i)$. This may be done by differentiating with respect to y the coefficient of x^n in the

purely formal expansion in x of $2^{-n+1}(x + x^2 + \cdots + x^{i-1} + yx^i + x^{i+1} + \cdots)^r$, setting $y = 1$ and summing over all values of r. We have therefore to evaluate

$$2^{-n+1} \cdot \sum_{r=2}^{n} r \binom{n - i - 1}{r - 2},$$

which is easily seen to give us the result

$$E(r_i) = (n - i + 3)2^{-i-1}, \qquad (i < n), \tag{2.8}$$

while it is obvious that

$$E(r_n) = 2^{-n+1}. \tag{2.9}$$

By use of similar devices the variances and covariances of the r_i may also be obtained. We omit the details of those calculations and also the presentation of the covariances, since the latter are not necessary for the proof of Theorem 2. The results are:

$$\sigma^2(r_i) = n\left(\frac{1}{2^{i+1}} + \frac{3 - 2i}{2^{2i+2}}\right) + \left(\frac{3 - i}{2^{i+1}} + \frac{3i^2 - 12i + 5}{2^{2i+2}}\right), \qquad (i < \tfrac{1}{2}n). \tag{2.10}$$

The limitation on the value of i is necessary because the processes for summing binomial coefficients with the aid of the device described above are no longer applicable. The matter is easily settled, however, for if X is a stochastic variable which can take only the values 0 or 1, then

$$\sigma^2(X) = E(X) - [E(X)]^2.$$

The r_i for $i > \frac{1}{2}n$ are such variables, so that

$$\sigma^2(r_i) = \frac{n - i + 3}{2^{i+1}} - \frac{(n - i + 3)^2}{2^{2i+2}}, \qquad (n > i > \tfrac{1}{2}n), \tag{2.11}$$

$$\sigma^2(r_n) = \frac{(2^{n-1} - 1)}{2^{2n-2}}. \tag{2.12}$$

Also without difficulty we have

$$\sigma^2(r_{\frac{1}{2}n}) = \frac{n + 6}{2^{\frac{1}{2}(n+4)}} - \frac{(n + 6)^2}{2^{n+4}} + \frac{1}{2^{n-2}}, \tag{2.13}$$

when n is even and > 2, and

$$E(r) = \tfrac{1}{2}(n + 1), \tag{2.14}$$

$$\sigma^2(r) = \tfrac{1}{4}(n - 1). \tag{2.15}$$

Finally,

$$E(r'_{(k+1)}) = (n - k + 1)2^{-k-1}. \tag{2.16}$$

The next results we shall need may be expressed in the following:

THEOREM 1: *As n approaches infinity, the joint distribution of the stochastic*

variables $\bar{r}_1, \cdots, \bar{r}_k, \bar{r}'_{(k+1)}$ (*k any fixed positive integer*), *approaches the multivariate normal distribution.*

This theorem is proved as follows: Make the substitutions

$$x_i = \frac{r_i - n\cdot 2^{-i-1}}{\sqrt{n}}, \qquad (i = 1, 2, \cdots, k),$$

$$x'_{(k+1)} = \frac{r'_{(k+1)} - n\cdot 2^{-k-1}}{\sqrt{n}}$$

in the expression

$$2^{-n+1}\frac{\left(\sum_{i=1}^{k} r_i + r'_{(k+1)}\right)!}{(r'_{(k+1)})!\prod_{i=1}^{k}(r_i)!}\binom{n - t - kr'_{(k+1)} - 1}{r'_{(k+1)} - 1},$$

which comes from (2.7), and regard t as equal to $\sum_{i=1}^{k} ir_i$. Replace the various factorials by their asymptotic approximations as given by Stirling's formula and simplify the resulting expression. The subsequent procedure is simple but laborious and we omit the details, which are like those of the classical proof of De Moivre's theorem as given, for example, in Frechet [1], p. 89.

We now prove the following theorem on additive partition functions:

THEOREM 2: *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions:*

(a). *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}, \tag{2.17}$$

(b). *the series*

$$\sum_{i=1}^{\infty} |f(i)|\, 2^{-\frac{1}{2}i}, \tag{2.18}$$

converges. Let $F(A)$, a function of the stochastic sequence A, be defined as follows:

$$F(A) = \sum_{i=1}^{s} f(a_i). \tag{2.19}$$

Then for any real y the probability of the inequality $\bar{F}(A) < y$, approaches

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

We restate this theorem without use of probabilistic terms:

Let A be any sequence of positive integers whose sum is a given integer n. Consider two sequences A to be different if they contain different elements or

the same elements arranged in a different order Let $f(x)$ and $F(A)$ be defined as above, with the aforementioned restrictions. Then there exist, for every positive integer n, two numbers E_n and σ_n, such that 2^{-n+1} multiplied by the number of sequences A for which the inequality

$$F(A) - E_n < y\sigma_n ,$$

holds, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

For convenience, the proof will be divided into a number of lemmas.

If $\varphi(y)$ is any continuous distribution function, then it is well known that $\varphi(y)$ is uniformly continuous and that consequently, for any arbitrarily small, positive ϵ, there exist two positive numbers, h and D, with the following properties

(a). If y_1 and y_2 are any real numbers such that $| y_1 - y_2 | < h$, then $| \varphi(y_1) - \varphi(y_2) | < \epsilon$,

(b) If y is such that $| y | > D$, then $\varphi(| y |) > 1 - \epsilon$, and $\varphi(-| y |) < \epsilon$

We now first prove

LEMMA 1 *Let X and Y be two stochastic variables, both of which possess finite means and variances Suppose that there exists a continuous distribution function $\varphi(y)$ and two small positive numbers ϵ and δ* (say $\epsilon < 1/10$, $\delta < 1/10$), *such that*

$$| P\{X < y\} - \varphi(y) | < \epsilon, \tag{2.20}$$

for all y, and

$$\frac{\sigma(Y)}{\sigma(X)} = \delta. \tag{2.21}$$

Let h and D be chosen as above for $\varphi(y)$, with the additional proviso that $h < \frac{1}{2}$ and $D > 1$ Suppose further that

$$\delta < \min\left(\frac{h}{4D}, \frac{h\epsilon}{8}\right). \tag{2.22}$$

Then

$$| P\{(\overline{X + Y}) < y\} - \varphi(y) | < 3\epsilon, \tag{2.23}$$

for all y.

PROOF. We have

$$\sigma^2(X + Y) = \sigma^2(X) + 2\sigma(XY) + \sigma^2(Y).$$

Since, as is well known,

$$| \sigma(XY) | \leq \sigma(X)\sigma(Y),$$

it follows from (2.21) that

$$\sigma(X + Y) = (1 + \delta')\sigma(X), \tag{2.24}$$

where $|\delta'| \leq \delta$. Hence

$$\sigma\left(\frac{Y - E(Y)}{\sigma(X + Y)}\right) < 2\delta. \tag{2.25}$$

From Tchebycheff's inequality and (2.21) it then follows that, if $d = h/4$,

$$P\left\{\left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} < 4\frac{\delta^2}{d^2}, \tag{2.26}$$

and

$$\frac{4\delta^2}{d^2} < \epsilon^2 < \epsilon. \tag{2.27}$$

Now

$$\begin{aligned}
P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d\right\} &= P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\
&\quad + P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} \\
&< P\{(\overline{X + Y}) < y\} + \epsilon \\
&= P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\
&\quad + P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} + \epsilon \\
&< P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y + d\right\} + 2\epsilon.
\end{aligned} \tag{2.28}$$

Hence, from (2.24)

$$\begin{aligned}
&P\{\bar{X} < (y - d)(1 + \delta')\} - \epsilon \\
&\qquad < P\{(\overline{X + Y}) < y\} < P\{\bar{X} < (y + d)(1 + \delta')\} + \epsilon
\end{aligned} \tag{2.29}$$

and consequently, from (2.20)

$$\begin{aligned}
&\varphi(y - d + y\delta' - d\delta') - 2\epsilon \\
&\qquad < P\{(\overline{X + Y}) < y\} < \varphi(y + d + y\delta' + d\delta') + 2\epsilon.
\end{aligned} \tag{2.30}$$

Now if $|y| \leq 2D$, then from (2.22)

$$d + |y\delta'| + d|\delta'| < \frac{h}{4} + \frac{h}{2} + \frac{h}{4} = h,$$

and if $|y| > 2D$, then also from (2 22)

$$|y| - d - |y\delta'| - d|\delta'| > |y|(1 - \delta) - \frac{h}{2} > \frac{3}{4}|y| > \frac{3}{2}D.$$

Recalling the definitions of h and D, it follows from (2.30) that, for all y,

$$\varphi(y) - 3\epsilon < P\{(\overline{X + Y}) < y\} < \varphi(y) + 3\epsilon. \tag{2 31}$$

This proves Lemma 1

LEMMA 2: *For any fixed pair a, b, of positive integers such that $a < b$,*

$$\lim_{n\to\infty} \frac{[E(r)]^{b-a}\,[E(r_b)]^a}{[E(r_a)]^b} = 1 \tag{2 32}$$

PROOF From (2.8), for fixed i

$$\frac{1}{n} E(r_i) \to 2^{-i-1},$$

and from (2.14) $\frac{1}{n} E(r) \to \frac{1}{2}$ as $n \to \infty$. The required result follows easily.

For any n we now define

$$B(k,n) = \sum_{i=1}^{k} r_i[f(i)],$$

and

$$C(k;n) = \sum_{i=k+1}^{n} r_i[f(i)].$$

Then

$$F(A) = B(k;n) + C(k,n).$$

LEMMA 3: *For any real y and any fixed positive integral k the probability that the stochastic variable $B(k;n)$ shall fulfill the inequality $\bar{B}(k,n) < y$ approaches $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\,dy$, as $n \to \infty$.*

PROOF. By Theorem 1, the stochastic variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k, \bar{r}'_{(k+1)'}$ are asymptotically jointly normally distributed. As an immediate consequence so are the variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k$, and hence $B(k;n)$, which is a linear function with constant coefficients $f(1), f(2), \cdots f(k)$, of $r_1, r_2, \cdots, r_k$, is asymptotically normally distributed

LEMMA 4. *There exists a constant $c > 0$, such that, for all n sufficiently large,*

$$\sigma^2(F(A)) > cn. \tag{2.33}$$

PROOF. For any sufficiently large, arbitrary, but fixed n, we will construct two sets, S_1 and S_2, of sequences A, with the following properties S_1 and S_2 have the same probability p, with p always greater than β, a fixed positive

constant which does not depend on n. Since the probabilities of S_1 and S_2 are equal, each possesses the same number of sequences A. Between the member sequences of the sets S_1 and S_2 we will establish a one-to-one correspondence such that, if A_1 is a member of S_1 and A_2 is its corresponding sequence in S_2, then

$$|F(A_1) - F(A_2)| > 2d\sqrt{n}, \tag{2.34}$$

where d is a fixed positive constant which does not depend on n.

It is easy to see that such a construction would prove the lemma. The probability of any sequence A is 2^{-n+1}. Hence the contribution of a corresponding pair A_1 and A_2 to the variance of $F(A)$ is by (2.34) not less than $2^{-n+2}d^2n$ and the contribution of the sets S_1 and S_2 is not less than $2\beta d^2 n$.

It remains then to carry out the construction of S_1 and S_2. For the sake of simplicity in notation, we shall carry out the construction with the assumption that the integers a and b of (2.17) are 1 and 2. It will be readily apparent, however, that the proof is perfectly general and with trivial changes holds for any pair a, b. This lemma is the only place where the hypothesis (2.17) is used. The latter condition is necessary because, if for every pair of positive integers i and j,

$$\frac{f(i)}{f(j)} = \frac{i}{j},$$

then $F(A)$ is a constant multiple of n, for $n = \sum_i ir_i$ and then

$$F(A) = \sum_i f(a_i) = \sum_i r_i f(i) = f(1) \sum_i ir_i = nf(1).$$

Each sequence A uniquely determines the "coordinate" complex

$$\{r_1, r_2, \cdots, r_n\}$$

which we prefer to write as the pair $L = (l, l')$:

$$l = \{r_1, r_2\},$$

$$l' = \{r_3, r_4, \cdots, r_n\}.$$

To each pair (l, l') there correspond in general many sequences A whose exact number may be explicitly given in terms of factorials. The totality of all A whose L have the same second member l' will be called the group determined by l', or just the group l'. The subset of a group l' all of whose A have the same r_1 will be called the family (l', r_1). All the A in the same family have the same L. For l' and r_1 determine r_2 through the equation $\sum_i ir_i = n$.

According to Theorem 1 for $k = 2$, r_1, r_2, r_3' are asymptotically jointly normally distributed. Let

$$\sigma_1 = \lim_{n\to\infty} \frac{\sigma(r_1)}{\sqrt{n}}$$

The limiting variances of r_2 and r_3' are constant multiples of $n\sigma_1^2$. Therefore the set H of all A whose L satisfy the constraints

$$\begin{aligned} \frac{n}{4} &< r_1 < \frac{n}{4} + \sqrt{n}\,\sigma_1 \\ \frac{n}{8} &< r_2 < \frac{n}{8} + \sqrt{n}\,\sigma_1 \\ \frac{n}{8} &< r_3' < \frac{n}{8} + \sqrt{n}\,\sigma_1 \end{aligned} \tag{2.35}$$

has, by virtue of the fact that the limiting correlation coefficients of the variables r_1, r_2, r_3' are all less than 1 in absolute value, a positive probability, which exceeds a fixed positive constant γ for sufficiently large n If any member sequence A of a family is in H, the entire family is obviously in H. Any sequence A belongs to one and only one family. Hence the set H may be decomposed in a disjunct way into entire families. Let $\left(l', \frac{n}{4} + h_1\right)$ be *any* family in H, where of course $0 < h_1 < \sqrt{n}\,\sigma_1$. Consider the (second) family $\left(l', \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)$. This family is not in H. We now wish to show that the probability of the second family exceeds c' times the probability of the first family, where c' is a fixed positive constant which does not depend on either n or the particular families in question

For the first family, let

$$r_1 = \frac{n}{4} + h_1, \qquad r_3' = \frac{n}{8} + h_3,$$

$$r_2 = \frac{n}{8} + h_2, \qquad r = \frac{n}{2} + h_1 + h_2 + h_3.$$

Hence

$$0 < h_i < \sqrt{n}\sigma_1 \qquad (i = 1, 2, 3). \tag{2.36}$$

For the second family we therefore have, since both families are in the same group,

$$\begin{aligned} r_1 &= \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1, \\ r_2 &= \frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2, \\ r_3' &= \frac{n}{8} + h_3, \\ r &= \frac{n}{2} + \sqrt{n}\,\sigma_1 + h_1 + h_2 + h_3. \end{aligned}$$

The ratio of the probability of the second family to that of the first family equals the ratio of the number of sequences A in the second family to the number of sequences A in the first family. By elementary combinatorics, since both families are in the same group, the latter ratio is

$$(2.37)\qquad \frac{\left(\frac{n}{2}+\sqrt{n}\,\sigma_1+h_1+h_2+h_3\right)!\quad \left(\frac{n}{4}+h_1\right)!\left(\frac{n}{8}+h_2\right)!}{\left(\frac{n}{4}+2\sqrt{n}\,\sigma_1+h_1\right)!\left(\frac{n}{8}-\sqrt{n}\,\sigma_1+h_2\right)!\left(\frac{n}{2}+h_1+h_2+h_3\right)!}$$

and hence exceeds

$$(2.38)\qquad \left(\frac{n}{2}+h_1+h_2+h_3\right)^{\sqrt{n}\,\sigma_1} \times\left(\frac{n}{4}+2\sqrt{n}\,\sigma_1+h_1\right)^{-2\sqrt{n}\,\sigma_1}\left(\frac{n}{8}-\sqrt{n}\,\sigma_1+h_2\right)^{\sqrt{n}\,\sigma_1}.$$

At this point, if we had been using the numbers a and b of (2.17), we would make use of Lemma 2. In the present case the result of that lemma is trivial. It is easy to see, therefore, that (2.38) equals

$$(2.39)\qquad \left(1+\frac{2h_1+2h_2+2h_3}{n}\right)^{\sqrt{n}\,\sigma_1} \times\left(1+\frac{8\sqrt{n}\,\sigma_1+4h_1}{n}\right)^{-2\sqrt{n}\,\sigma_1}\left(1-\frac{8\sqrt{n}\,\sigma_1-8h_2}{n}\right)^{\sqrt{n}\,\sigma_1},$$

which, in view of (2.36), exceeds

$$(2.40)\qquad \left(1+\frac{12\sigma_1}{\sqrt{n}}\right)^{-2\sqrt{n}\,\sigma_1}\cdot\left(1-\frac{8\sigma_1}{\sqrt{n}}\right)^{\sqrt{n}\,\sigma_1}$$

which, in turn, for sufficiently large n, exceeds

$$(2.41)\qquad \tfrac{1}{2}\cdot e^{-24\sigma_1^2-8\sigma_1^2}=\tfrac{1}{2}e^{-32\sigma_1^2}=c'.$$

We are now ready to construct S_1 and S_2. Let

$$f_1=(l',r_1)$$

be any family in H and consider the family

$$f_2=(l',r_1+2\sqrt{n}\,\sigma_1).$$

Select in any manner whatsoever $c'\nu$ of the sequences A in f_1, where ν is the total number of sequences in f_1. Call this set of sequences f^*. Select in any manner whatsoever $c'\nu$ sequences from f_2 and call this set f^{**}. That there exist at least $c'\nu$ sequences in f_2 is assured by equation (2.41). In any manner whatsoever establish a one-to-one correspondence between the sequences of f^* and f^{**}. Suppose A_1 and A_2 are corresponding sequences. Since f^* and f^{**} belong to the same group, and since $f(2)\neq 2f(1)$, we have

$$\begin{aligned}(2.42)\qquad |F(A_1) - F(A_2)| &= |f(2)\sqrt{n}\,\sigma_1 - 2f(1)\sqrt{n}\,\sigma_1| \\ &= |f(2) - 2f(1)|\sqrt{n}\,\sigma_1,\end{aligned}$$

so that (2.34) holds with

$$(2.43)\qquad d = \tfrac{1}{4}\,|f(2) - 2f(1)|\,\sigma_1 .$$

Now proceed in this manner for all the families f_1 in H. The union of all the sets f^* is the set S_1 and the union of all the sets f^{**} is the set S_2 It is clear that, since the probability of H exceeds γ, the probability p of S_1 exceeds $\beta = c'\gamma$. This proves Lemma 4

LEMMA 5. *For any arbitrarily small positive number ξ there exists a positive integer $\mu(\xi)$, such that for any $k > \mu(\xi)$ and all n greater than a fixed lower bound,*

$$(2.44)\qquad \sigma^2[C(k;n)] < \xi n.$$

PROOF Since

$$C(k;n) = \sum_{i=k+1}^{n} r_i f(i),$$

and, as is well known,

$$|\sigma(XY)| \leq \sigma(X)\sigma(Y)$$

we have

$$(2.45)\qquad \sigma^2[C(k;n)] \leq \left[\sum_{i=k+1}^{n} |f(i)|\,\sigma(r_i)\right]^2$$

From (2.10) it follows readily that

$$(2.46)\qquad \sigma^2(r_i) < \frac{n}{2^i} + \frac{5}{2^{i+1}} + \left(\frac{-i}{2^{i+1}} + \frac{3i^2}{2^{2i+2}}\right),$$

and the quantity in parentheses in the right member of (2.46) is easily seen to be negative, so that, for $i < \frac{1}{2}n$ and $n \geq 3$,

$$(2.47)\qquad \sigma(r_i) < \sqrt{2n}\,2^{-\frac{1}{2}i}$$

From (2.11) and the definition of r_i, it follows easily that (2.47) holds also when $i > \frac{1}{2}n$ and $n \geq 3$

Hence, in view of (2.12), (2.13), and the convergence of the series in (2.18), the desired result follows from (2.45).

LEMMA 6. *Let the ξ of Lemma 5 be $< \frac{1}{4}c$, where c is as in Lemma 4. Then for $k > \mu(\xi)$ and n larger than a fixed lower bound*

$$(2.48)\qquad \sigma^2(B(k;n)) > \tfrac{1}{4}cn$$

PROOF: Since

$$F(A) = B(k,n) + C(k;n),$$

we have

$$\sigma^2(F(A)) = \sigma^2(B(k;n)) + \sigma^2(C(k;n)) + 2\sigma(BC)$$
$$\leq \sigma^2(B) + \sigma^2(C) + 2\sigma(B)\sigma(C) = (\sigma(B) + \sigma(C))^2.$$

Hence from (2.33) and (2.44) $\sqrt{cn} < \sigma(B) + \frac{1}{2}\sqrt{cn}$ and the required result follows.

Proof of the Theorem: Let ϵ be an arbitrarily small positive number. For all n sufficiently large we have, by Lemma 3,

$$\left| P\{\bar{B}(k;n) < y\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy \right| < \epsilon,$$

for all y. For a small ξ to be chosen later and large enough k and n we have, by Lemmas 5 and 6,

$$\frac{\sigma(C(k;n))}{\sigma(B(k,n))} = \delta < \frac{4\xi}{c}. \tag{2.49}$$

Now let the $\varphi(y)$ of Lemma 1 be defined as

$$\varphi(y) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

and choose h and D as in Lemma 1 for our present ϵ. Since c is fixed and ξ still at our disposal, choose ξ sufficiently small so that the δ of (2.49) satisfies (2.22) Since the hypothesis of Lemma 1 is satisfied, we have, from (2.23) and Lemma 3, for all n sufficiently large,

$$| P\{\bar{F}(A) < y\} - \varphi(y) | < 3\epsilon$$

for all y This is the required result.

3. Partitions of two integers. Let n_1 and n_2 be positive integers, $n_1 + n_2 = n$. $\frac{n_1}{n} = e_1$, $\frac{n_2}{n} = e_2$, and $e = \max(e_1, e_2)$. Let $V = (v_1, v_2, \cdots, v_s)$ be any sequence of positive integers v_i $(i = 1, 2, \cdots, s)$ where $a_1 + a_3 + a_5 + \cdots$ equals either one of n_1 and n_2, while $a_2 + a_4 + a_6 + \cdots$ equals the other. Such sequences are of statistical importance (cf. Wald and Wolfowitz [2]). As before, sequences V with different elements or with the same elements in different order will be considered different and to each sequence V will be assigned the same probability, which is therefore easily seen to be $\frac{n_1!\, n_2!}{n!}$.

Let r_{1i} be the number of elements equal to i in that one of the two sequences $(a_1, a_3, a_5, \cdots)$ and $(a_2, a_4, a_6, \cdots)$ the sum of whose elements is n_1 and let r_{2i} be the corresponding number for the other sequence. Let

$$s_i = r_{1i} + r_{2i},$$

$$r_1 = \sum_i r_{1i}, \qquad r_2 = \sum_i r_{2i},$$

$$s = r_1 + r_2, \qquad r'_{1(k+1)} = \sum_{i=k+1}^{n_1} r_{1i}$$

$$r'_{2(k+1)} = \sum_{i=k+1}^{n_2} r_{2i}.$$

The necessary computations such as are given in the beginning of the previous section have been performed by Mood [3] and we summarize them as follows:

THEOREM 3 (Mood). *As n approaches infinity while e_1 and e_2 remain constant, the joint distribution of the stochastic variables*

$$\bar{r}_{11}, \bar{r}_{12}, \cdots, \bar{r}_{1k}, \bar{r}'_{1(k+1)}, \bar{r}_{21}, \bar{r}_{22}, \cdots \bar{r}_{2k}$$

(where k is any fixed positive integer), approaches the multivariate normal distribution.

Mood (loc. cit) gives the following parameters, with the convention that

$$x^{(i)} = x(x-1)(x-2)\cdots(x-i+1): \tag{3.1}$$

$$E(r_{1i}) = \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}, \tag{3.2}$$

$$\lim_{n\to\infty} \frac{E(r_{1i})}{n} = e_1^i e_2^2, \tag{3.3}$$

$$\lim_{n\to\infty} \frac{E(r'_{1(k+1)})}{n} = e_1^{k+1} e_2, \tag{3.4}$$

$$\sigma^2(r_{1i}) = \frac{n_2^{(2)}(n_2+1)^{(2)} n_1^{(2i)}}{n^{(2i+2)}} + \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\left(1 - \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\right), \tag{3.5}$$

$$\lim_{n\to\infty} \frac{\sigma^2(r_{1i})}{n} = e_1^{2i-1} e_2^3[(i+1)^2 e_1 e_2 - i^2 e_2 - 2e_1] + e_1^i e_2^2. \tag{3.6}$$

The corresponding parameters for r_{2i} may be obtained from the above by interchange of n_1 and n_2. Also

$$\lim_{n\to\infty} \frac{E(r_1)}{n} = \lim_{n\to\infty} \frac{E(r_2)}{n} = e_1 e_2. \tag{3.7}$$

For additive partition functions we have the following theorem:

THEOREM 4 *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions:*

a) *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}; \tag{3.8}$$

b) *the series*

$$\sum_{i=1}^{\infty} | f(i) | e^{i/2} \tag{3.9}$$

converges. Let $F(V)$*, a function of the stochastic sequence* V*, be defined as follows:*

$$F(V) = \sum_{i=1}^{s} f(v_i) \tag{3.10}$$

Then for any real y *the probability of the inequality* $F(V) < y$ *approaches*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$*, while* e_1 *and* e_2 *remain constant.*

The basic idea of the proof of this theorem is the same as that of the proof of Theorem 2. We omit all the steps which can be written without difficulty by analogy to those in Theorem 2 and present only those where some major change is necessary. The numbering of the lemmas will correspond to that of Theorem 2.

LEMMA 2. *For any fixed pair,* a *and* b*, of positive integers such that* $a < b$*,*

$$[E(r_1)]^{b-a} \cdot [E(r_2)]^{b-a} \cdot [E(r_{1b})]^{a} \cdot [E(r_{2b})]^{a} \cdot [E(r_{1a})]^{-b} \cdot [E(r_{2a})]^{-b} \to 1, \tag{3.11}$$

as $n \to \infty$.

The proof is the same as before.

The following are the definitions corresponding to those of Theorem 2:

$$B(k;n) = \sum_{i=1}^{k} s_i f(i),$$

$$C(k;n) = \sum_{i=k+1}^{n} s_i f(i).$$

Then as before

$$F(V) = B(k;n) + C(k;n).$$

LEMMA 4. Statement is the same as that for Theorem 2. The following important changes must be made in the proof:

Each sequence V determines the coordinate complex

$$\begin{Bmatrix} r_{11}, r_{12}, \cdots, r_{1n} \\ r_{21}, r_{22}, \cdots, r_{2n} \end{Bmatrix}$$

also

$$l = \begin{Bmatrix} r_{11}, r_{12} \\ r_{21}, r_{22} \end{Bmatrix},$$

and

$$l' = \begin{Bmatrix} r_{13}, \cdot & , r_{1n} \\ r_{23}, \cdots, r_{2n} \end{Bmatrix}.$$

The set H is the set of all V whose L satisfy the constraints

$$ne_1e_2^2 < r_{11} < ne_1e_2^2 + \sqrt{n}\,\sigma_{11},$$

$$ne_1^2e_2^2 < r_{12} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11},$$

$$ne_1^2e_2 < r_{21} < ne_1^2e_2 + \sqrt{n}\,\sigma_{11},$$

$$ne_1^2e_2^2 < r_{22} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11},$$

$$ne_1^3e_2 < r'_{13} < ne_1^3e_2 + \sqrt{n}\,\sigma_{11},$$

where

$$\sigma_{11} = \lim_{n\to\infty} \frac{\sigma(r_{11})}{\sqrt{n}}$$

The representative family for H is characterized by

$$(l', ne_1e_2^2 + h_{11}),$$

and this family is compared with the family

$$(l', ne_1e_2^2 + 2\sqrt{n}\,\sigma_{11} + h_{11}).$$

For the members of the family in H

$$r_{11} = ne_1e_2^2 + h_{11} = nm_{11} + h_{11},$$

$$r_{12} = ne_1^2e_2^2 + h_{12} = nm_{12} + h_{12},$$

$$r_{21} = ne_1^2e_2 + h_{21} = nm_{21} + h_{21},$$

$$r_{22} = ne_1^2e_2^2 + h_{22} = nm_{22} + h_{22},$$

$$r'_{13} = ne_1^3e_2 + h_{13} = nm'_{13} + h_{13},$$

$$r_1 = ne_1e_2 + h' = nm + h,$$

$$|r_2 - r_1| \leq 1,$$

where

(3 12) $$h_{ij} < \sqrt{n}\,\sigma_{11},$$

(3.13) $$h = h_{11} + h_{12} + h_{13}$$

And for the members of the second family

$$r_{11} = nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11},$$

$$r_{12} = nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12},$$

$$
\begin{aligned}
r_{21} &= nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21}\,, \\
r_{22} &= nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22}\,, \\
r'_{13} &= nm'_{13} + h_{13}\,, \\
r_1 &= nm + \sqrt{n}\,\sigma_{11} + h, \\
&\quad | r_2 - r_1 | \leq 1,
\end{aligned}
$$

with

$$| \theta_{21} | \leq 1, \qquad | \theta_{22} | \leq 1$$

To the expression (2.37) corresponds the expression (3.14), with $| \theta | \leq 1$:

$$
\begin{aligned}
&\frac{(nm_{11} + h_{11})!\,(nm_{12} + h_{12})!}{(nm + h)!} \times \frac{(nm_{21} + h_{21})!\,(nm_{22} + h_{22})!}{(nm + h)!} \\
&\times \frac{(nm + h + \sqrt{n}\,\sigma_{11})!}{(nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})!\,(nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})!} \\
&\times \frac{(nm + h + \sqrt{n}\,\sigma_{11} + \theta)!}{(nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21})!\,(nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22})!},
\end{aligned}
\tag{3.14}
$$

which exceeds

$$
\begin{aligned}
(nm + h)^{2\sqrt{n}\,\sigma_{11}} &\times (nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})^{-2\sqrt{n}\,\sigma_{11}} \\
&\times (nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})^{\sqrt{n}\,\sigma_{11}} \\
&\times (nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21})^{-2\sqrt{n}\,\sigma_{11}} \\
&\times (nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22})^{\sqrt{n}\,\sigma_{11}}
\end{aligned}
\tag{3.15}
$$

Employing Lemma 2, we find that (3.15) equals

$$
\begin{aligned}
\left(1 + \frac{h}{nm}\right)^{2\sqrt{n}\,\sigma_{11}} &\times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\
&\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{12}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\
&\times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{21}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\
&\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{22}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}}.
\end{aligned}
\tag{3.16}
$$

In view (3.12) and (3.13), (3.16) exceeds

$$
\begin{aligned}
&\left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\
&\qquad \times \left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}},
\end{aligned}
\tag{3.17}
$$

which, for sufficiently large n, in turn exceeds

$$\tfrac{1}{2}\, e^{-\sigma^2_{11}} \left(\frac{6}{m_{11}} + \frac{1}{m_{12}} + \frac{6}{m_{21}} + \frac{1}{m_{22}}\right) = c'. \tag{3.18}$$

LEMMA 5 Statement is the same as for Theorem 2 The proof then proceeds as follows·

We have

$$\sigma^2(C(k,n)) \leq \left(\sum_{i=1}^{2} \sum_{j=k+1}^{n} | f(j) | \sigma(r_{ij})\right)^2. \tag{3.19}$$

From an examination of (3 5) and (3.6) we may see without any difficulty that the second of the three terms of the right member of (3.5) (after removal of parentheses) is asymptotically equal to n times the last term of the right member of (3 6) and hence that the other two terms of the right member of (3 5) are asymptotically equal to n times the right member of (3 6) without its last term. Now when

$$\frac{i+1}{i} \sqrt{e_1} < 1$$

which will always occur when i is equal to or greater than a sufficiently large fixed integer μ, that part of the right member of (3.6) which is in square brackets is easily seen to be negative. Hence from the definition of asymptotic equivalence it follows that, for all n sufficiently large,

$$\frac{n_2^{(2)}(n_2+1)^{(2)}\, n_1^{(2\mu)}}{n^{(2\mu+2)}} < \frac{(n_2+1)^{(2)}(n_2+1)^{(2)}\, n_1^{(\mu)}\, n_1^{(\mu)}}{n^{(\mu+1)}\, n^{(\mu+1)}}, \tag{3.20}$$

and

$$\frac{(n_2+1)^{(2)}\, n_1^{(\mu)}}{n^{(\mu+1)}} < 2ne^{\mu+2} < 2ne^{\mu} \tag{3.21}$$

Hence, for all n sufficiently large,

$$\sigma^2(r_{1\mu}) < 2ne^{\mu} \tag{3.22}$$

Now consider the expression (3.5) for $i = \mu$ and $i = \mu + 1$ Passage from μ to $\mu + 1$ multiplies the first term of the right member of (3 5) by

$$\frac{(n_1 - 2\mu)(n_1 - 2\mu - 1)}{(n - 2\mu - 2)(n - 2\mu - 3)}, \tag{3.23}$$

and the third term of the right member by

$$\frac{(n_1 - \mu)^2}{(n - \mu - 1)^2} \tag{3.24}$$

It is easy to see that for large but fixed μ and all n greater than a lower bound which is a function of μ only, the expression (3 23) is less than the expression (3.24). Hence, in view of (3.20), the sum of the first and third terms of the

right member of (3.5) for $i = \mu + 1$ is negative. Now consider what happens to the second term of the right member of (3.5) when i goes from μ to $\mu + 1$. It is multiplied by

$$\frac{(n_1 - \mu)}{(n - \mu - 1)}, \tag{3.25}$$

which, also for large but fixed μ and all n larger than a lower bound which is a function of μ only, is easily seen to be less than e. Consequently

$$\sigma^2(r_{1(\mu+1)}) < 2ne^{\mu+1}. \tag{3.26}$$

It can be seen without difficulty that such a passage of (3.5) to the next higher index is always accompanied by multiplication by expressions similar to (3.23), (3.24), and (3.25), for which similar inequalities hold and that consequently

$$0 \leq \sigma^2(r_{1i}) < 2ne^i, \tag{3.27}$$

and for similar reasons

$$0 \leq \sigma^2(r_{2i}) < 2ne^i,$$

for all i not less than μ and for all n greater than a lower bound which is a function of μ only (although it may be necessary to increase the original μ so that *both* the last two equations hold). The required result follows from (3.19) and the convergence of the series (3.9).

The proof of Theorem 4 follows along the same lines as that of Theorem 2.

When $f(x) \equiv 1$, $F(V) \equiv U(V)$, the statistic discussed in [2]. Other such results follow from specialization of $f(x)$. Theorem 4 may also be generalized so that the elements v_i which add up to n_1 are operated on by a function f_1, while the elements v_i which add up to n_2 are operated on by another function f_2, but this is easy to see and we do not go into the details.

4. Tests of hypotheses in the non-parametric case. The great advances that have been made in mathematical statistics in recent years have been in two directions. On the one hand, the foundations of statistics, the theory of estimation and of testing hypotheses have been put on a rigorous basis of probability theory, and on the other, powerful methods for obtaining critical regions and confidence intervals and criteria for appraising their efficacy have been developed. Most of these developments have this feature in common, that the distribution functions of the various stochastic variables which enter into their problems are assumed to be of known functional form, and the theories of estimation and of testing hypotheses are theories of estimation of and of testing hypotheses about, one or more parameters, finite in number, the knowledge of which would completely determine the various distribution functions involved. We shall refer to this situation for brevity as the parametric case, and denote the opposite situation, where the functional forms of the distributions are unknown, as the non-parametric case.

The literature of theoretical statistics, therefore, deals principally with the parametric case The reasons for this are perhaps partly historic, and partly the fact that interesting results could more readily be expected to follow from the assumption of normality. Another reason is that, while the parametric case was for long developed on an intuitive basis, progress in the non-parametric case requires the use of modern notions. However, the needs of theoretical completeness and of practical research require the development of the theory of the non-parametric case. The purpose of the following section is to contribute to this theory.

Brief mention of some of the literature may be made here The problem of parametric estimation by confidence intervals, was put on a rigorous foundation by Neyman [4] and extended to the estimation of distribution functions in the non-parametric case by means of confidence belts by Wald and Wolfowitz [5]. Problems of testing non-parametric hypotheses have been treated in various places. The rank correlation coefficient has been used for a long time to test the independence of two variates. Its distribution was shown to be asymptotically normal by Hotelling and Pabst [6] and its small sample distribution was discussed by Olds [7]. The problem of two samples has been discussed, among others, by Thompson [8], Dixon [9] and Wald and Wolfowitz [2]. In 1937, Friedman [10] posed the non-parametric analogue of the problem in the analysis of variance and proposed a very ingenious solution.

All these proposed solutions have this in common, that there exists no general principle which can be applied in each particular case to obtain a critical region, a role which is performed in the parametric case by Fisher's principle of maximum likelihood and the likelihood ratio criterion (Neyman and Pearson, [11]), whose validity, at least for large samples, has been established by Wald ([12], [13]). In each problem the solutions proposed have been intuitive and usually based on an analogy to the corresponding problem in the parametric case. Thus the principal justification for the use of the rank correlation coefficient is that its distribution is independent of the unknown distribution function (under the null hypothesis) and that its structure resembles that of the ordinary correlation coefficient. But any function of the order relations among the variates (cf. [2], p. 148) has a distribution which is independent of the unknown population distribution under the null hypothesis The same objection may be made to papers [8], [9], [10], [2], except that in [2], although the solution there proposed is an intuitive one, the criterion of consistency is extended from the parametric case to the non-parametric one. The fulfilment of this condition is a minimal requirement of a good test and on this basis the solution proposed in one of the previous papers cannot be considered a good one.

In the following section we shall show that the likelihood ratio criterion may be extended to the non-parametric case where the test must be made on the order relations among the observations and that for a certain class of these problems which fulfill the same requirement as that for the application of the likelihood ratio criterion in the parametric case it would thus appear to furnish

a general method by which statistics may be obtained for a specific problem. We shall show this by applying it to the problem of two samples This will serve to explain the method. Another problem will be discussed later. The ultimate justification of any statistic must be its power function, which ought therefore to constitute the next subject of investigation for these problems. Since for problems in the non-parametric case it is almost certain that uniformly most powerful tests do not exist, the question of determining the alternatives with respect to which proposed tests are powerful is particularly important.

5. The problem of two samples. Let X and Y be two stochastic variables with the distribution functions $f(x)$ and $g(x)$, respectively. (The term distribution function will always denote the cumulative distribution function. The letter P followed by an expression in braces will stand for the probability of the relation in braces. Hence $P\{X < x\} = f(x)$ for all x.) $f(x)$ and $g(x)$ are assumed continuous. The n_1 observations $x_1, x_2 \cdots, x_{n_1}$ and n_2 observations $y_1, y_2, \cdots, y_{n_2}$ are made on X and Y respectively. The (null) hypothesis to be tested is that $f(x) \equiv g(x)$. The admissible alternatives are all continuous distribution functions $f(x)$ and $g(x)$ such that $f(x) \not\equiv g(x)$. The $n_1 + n_2 = n$ observations are arranged in ascending order of size, thus: $Z = z_1, \cdots, z_n$ where $z_1 < z_2 < \cdots < z_n$ (the probability that $z_i = z_{i+1}$ is 0). Let $V = v_1, v_2, \cdots, v_n$ be a sequence defined as follows: $v_i = 0$ if z_i is a member of the set $x_1, x_2, \cdots, x_{n_1}$ and $v_i = 1$ if z_i is a member of the set $y_1, y_2, \cdots, y_{n_2}$. Then any statistic used to test the null hypothesis must be a function only of V ([2], p. 148).

We now apply the method of Neyman and Pearson [11] as follows. Ω is the totality of all couples $(d_1(x), d_2(x))$ of continuous distribution functions. The set ω, a subset of Ω, is the totality of all couples of distribution functions for which $d_1 \equiv d_2$. The sample space is the totality of all sequences V. The null hypothesis states that (f, g) is a member of ω. The admissible alternatives are that (f, g) is a member of Ω not in ω. The distribution of any function of V is the same for all members of ω. Hence this essential requirement on the statistic to be selected for the application of the likelihood ratio criterion (cf. [11]) is satisfied by any statistic which is a function of V alone. Furthermore, all sequences V have the same probability if the null hypothesis is true ([2], p. 149) The numerator of the likelihood ratio is therefore a function only of n_1 and n_2, is the same for all V, and is therefore of no further interest. Hence $T'(V)$, a function of V which is a monotonic function of the likelihood ratio for this problem, may be defined as the denominator of the likelihood ratio, as follows: Let $P\{V; (d_1, d_2)\}$ be the probability of V when $f \equiv d_1$, and $g \equiv d_2$. Then

$$T'(V) = \max_{\Omega} P\{V; (d_1, d_2)\}.$$

The critical values of $T'(V)$ are the large values. However, we may use instead of $T'(V)$ a convenient monotonic function of $T'(V)$.

As an approximation to $T'(V)$ we propose $T(V)$, a statistic which is obtained on the assumption that for a given V a couple (d_1^*, d_2^*) which is essentially the same as that of the two sample distribution functions corresponding to the particular V approximates a couple which maximizes the right member of (5.1). (We say "*a*" couple because it cannot be unique.) This assumption seems a reasonable one, particularly for large samples. Only the form of (d_1^*, d_2^*) is assumed and the missing parameters are obtained in accordance with (5.1). Before describing the matter precisely, it must be stressed that this is offered only as a plausible approximation. For certain extreme V, for example, like those where zeros and ones nearly alternate, this is definitely not the maximizing couple In spite of this the statistic $T(V)$ assigns to these V values which are furthest removed from the critical region for any level of significance, as indeed any good statistic should

We first define a "*run*" as in [2], p 149 A subsequence $v_{(t+1)}, v_{(t+2)}, \cdots, v_{(t+r)}$ of V (where r may also be 1) is called a "*run*" if $v_{(t+1)} = v_{(t+2)} = \cdots = v_{(t+r)}$ and if $v_t \neq v_{(t+1)}$ when $t > 0$ and if $v_{(t+r)} \neq v_{(t+r+1)}$ when $t + r < n$ Let l_{1j} be the number of elements in the j^{th} *run* of elements 0, and l_{2j} the number of elements in the j^{th} *run* of elements 1 Suppose for a moment that the first element in V is a 0. Consider the following situation. There is an interval $[a_1, a_2]$, $a_1 < a_2$, on the line $-\infty < x < +\infty$ such that

$$P\{a_1 \leq X \leq a_2\} > 0, \qquad P\{a_1 \leq Y \leq a_2\} = 0,$$
$$P\{X < a_1\} = P\{Y < a_1\} = 0.$$

This is followed by an interval $[b_1, b_2]$, $b_1 = a_2$, such that $P\{b_1 \leq X \leq b_2\} = 0$, $P\{b_1 \leq Y \leq b_2\} > 0$ This is in turn followed by an interval $[a_3, a_4]$, $a_3 = b_2$, such that $P\{a_3 \leq X \leq a_4\} > 0$, $P\{a_3 \leq Y \leq a_4\} = 0$, etc. It is clear that the lengths and location of the intervals described are immaterial, provided only that they do not overlap. Also the distributions of X and Y within each interval are immaterial, provided only that they are continuous All that matters for finding $P\{V, (d_1^*, d_2^*)\}$ is that the number and the order of the disjunct intervals shall be the same as those of the runs in V, (i.e, intervals of positive probability for X must alternate with intervals of positive probability for Y, the number of intervals of positive probability for X and for Y must equal respectively the number of runs of the element 0 and the number of runs of the element 1, and the probability of the first interval on the left shall be positive for X or for Y according as the first run in V is of elements 0 or of elements 1, with the same relation obtaining between the last interval on the right and the last run in V) and the probability of these intervals. Let P_{1j} be the sought for probability of the interval which corresponds to the jth run of elements 0 and P_{2j} the probability of the interval which corresponds to the jth run of elements 1. In order to obtain V, it is necessary that the elements constituting each run shall fall into its corresponding interval. Then clearly by the multinomial theorem

$$P\{V; (d_1^*, d_2^*)\} = \prod_i n_i! \left(\prod_j (l_{ij}!)^{-1} P_{ij}^{l_{ij}}\right) \tag{5.2}$$

where $i = 1, 2$ and where, when i is fixed, the product with respect to j is taken over all runs of the corresponding element. The right member of (5.2) is to be maximized with respect to the P_{ij}, subject of course to the constraints

$$\sum_j P_{ij} = 1 \qquad (i = 1, 2). \tag{5.3}$$

Then it may easily be verified that the maximum occurs when

$$P_{ij} = \frac{l_{ij}}{n_i} \qquad (i = 1, 2) \tag{5.4}$$

For, after multiplying by a constant and taking the logarithm we introduce two Lagrange multipliers μ_1 and μ_2 so that the maximizing P_{ij} are given by the equations (5.3) and those obtained by equating to zero all the partial derivatives of

$$\sum_i \sum_j (l_{ij} \log P_{ij} - \mu_i P_{ij}).$$

The latter are therefore

$$\mu_i = \frac{l_{ij}}{P_{ij}} \qquad (i = 1, 2),$$

for all j, whence (5.4) follows. It is easy to see that the extremum thus obtained is a maximum and also an absolute maximum. The sought-for statistic $T(V)$ is then the right member of (5.2) after the results (5.4) have been inserted. It may be simplified by removing all factors which are functions only of n_1 and n_2 (since these will then be the same for all V) and recalling that

$$\sum_j l_{ij} = n_i \qquad (i = 1, 2). \tag{5.5}$$

It will be convenient to take the logarithm of the resulting expression, so that with a slight change of notation we finally have

$$T(V) = \sum_i \sum_j \bar{l}_{ij} \tag{5.6}$$

where

$$\bar{l}_{ij} = \log\left(\frac{l_{ij}^{l_{ij}}}{l_{ij}!}\right). \tag{5.7}$$

This result is immediately extensible to the problem of k samples and by way of summary we recapitulate it as follows:

Let there be given k stochastic variables $X_1, \cdots, X_k$ with the respective distribution functions $f_1(x), \cdots, f_k(x)$, about which nothing is known except that they are continuous. Random independent observations, n_i in number, are made on X_i $(i = 1, \cdots, k)$. It is desired to test the hypothesis that $f_1 \equiv f_2 \equiv \cdots \equiv f_k$, the admissible alternatives being all k-tuples of continuous distribution functions. The sequence V is obtained from the sequence Z by

replacing an observation on X_i by the element i. Let l_{ij} be the number of elements in the jth run of elements i. Then the corresponding statistic for testing the null hypothesis is $T_k(V)$ or any monotonic function of it, where

$$T_k(V) = \sum_{i=1}^{k} \sum_{j} \bar{l}_{ij}$$

and $\bar{l}_{ij}$ is given by (5.7). The large values of $T_k(V)$ are the critical values.

Let r_{ij} denote the number of runs of length j in the elements i. Let $\sum_j r_{ij} = r_i$. Of course $\sum_j jr_{ij} = n_i$. Also let $s_j = \sum_i r_{ij}$. Then

$$T_k(V) = \sum_i \sum_j \bar{j} r_{ij} \tag{5.8}$$

and

$$T_k(V) = \sum_j \bar{j} s_j. \tag{5.9}$$

If a table were constructed of the numbers (5.7) from 1 to 50, say, or from 1 to 100, this would cover most of the cases arising in practice. The calculation of $T_k(V)$ by means of (5.9) would then be so simple that it could be performed very expeditiously by an ordinary clerk and with very much less labor than is required for most statistics in common use, like the correlation coefficient, for example. As a matter of interest we note that

$$\bar{1} = 0$$

$$\bar{2} = .693$$

$$\bar{3} = 1.50$$

$$\bar{4} = 2.37$$

$$\bar{5} = 3.26$$

and that

$$\bar{p} < p \tag{5.10}$$

where p is any integer ≥ 1. (5.10) follows from the fact that

$$p! > (\sqrt{2\pi p} - 1)p^p e^{-p}.$$

The distribution of $T(V)$ may be found for small samples by enumerating the sequences V, all of which have the same probability under the null hypothesis, and assigning to each V its $T(V)$. The critical region consists of the V's for which $T(V)$ takes the largest values, taken in sufficient number to make the critical region of proper size. It will not be necessary to enumerate all the V's, since it is readily apparent that certain V's can never belong to a critical region of any reasonable size. (Roughly speaking, a V with a large number of runs of short length will yield a small $T(V)$ and vice versa.) For large samples, the result of Section 3 is available, with $f(x) = \bar{x}$. From (5.10) it follows

easily that the corresponding series (3.9) is convergent, so that $\overline{T}(V)$ is asymptotically normally distributed. It must be remembered when using tables of the normal distribution that the critical region of $\overline{T}(V)$ lies in only one "tail" of the normal curve. The greatest difficulty will occur for samples of moderate size. Methods like those of Olds [7] will probably help there. It is highly unlikely that any practicable formula which would give the exact distribution of $T(V)$ exists.

A few brief remarks may be made here on a related problem. Suppose we have observations from two bivariate populations about the distributions of both of which nothing is known except that they are continuous and it is sought to test whether the two populations have the same distribution functions. Suppose further that it were required that the statistic used for this purpose be invariant under any topologic transformation of the whole plane into itself. At this point we quote the following topologic theorem, the proof of which was communicated to the author by Dr. Herbert Robbins: *Let x_1, y_1, x_2, y_2, $\cdots$, x_p, y_p be any $2p$ distinct points in the plane. There exists a topologic transformation of the whole plane into itself which takes x_i into y_i $(i = 1, 2, \cdots, p)$.* As a consequence of this theorem we get the absurd result that the required statistic must be a constant. Hence this statistical problem can have no solution.

As a matter of interest this statistical problem would have no solution even if it were not for the topologic theorem. The fact is that a continuous distribution on a line remains continuous under a topologic transformation of the whole line into itself, but a continuous distribution in a k-dimensional (Euclidean) space $(k > 1)$ may become discontinuous under a topologic transformation of the whole space into itself. (The probability distribution in the first space always determines a probability distribution in the transformed space, for probability functions are defined over all Borel sets of the space (cf. [15], p. 7) and a topologic transformation carries Borel sets into Borel sets (cf. [16], p. 195, Theorem II)). Consider the following example in the plane: A bivariate distribution function assigns probability 1 to a line L oblique to the coordinate axes, while any interval which contains no segment of the line L has probability 0. On the line L the (one-dimensional) probability distribution may be arbitrary, provided it is continuous. The bivariate distribution function is without difficulty seen to be continuous. Now rotate the coördinate axes until one of them is parallel to L. It is easy to see that after the rotation the bivariate distribution function is discontinuous.

The question of whether a useful statistical problem could be obtained by properly delimiting the class of transformations which are to leave the statistic invariant and the solution of such a problem remain to be investigated.

6. The problem of the independence of several variates. This is an important practical problem and one of the earliest discussed in the literature (cf., for example, [6]). Let X_1 and X_2 be stochastic variables with the joint (cumulative) distribution function $F(x_1, x_2)$ which is known to be continuous in both variables

jointly (i.e., $F(x_1, x_2) = P\{X_1 < x_1 ; X_2 < x_2\}$, where the right member is the probability of the occurrence of *both* the relations in braces). The marginal distributions $f_1(x_1)$ and $f_2(x_2)$ of X_1 and X_2 respectively are defined as follows:

$$f_1(x_1) = P\{X_1 < x_1\} = \lim_{x_2 \to +\infty} F(x_1, x_2),$$

$$f_2(x_2) = P\{X_2 < x_2\} = \lim_{x_1 \to +\infty} F(x_1, x_2).$$

(It is easy to see that the continuity of $F(x_1, x_2)$ implies the continuity of $f_1(x_1)$ and $f_2(x_2)$.)

The n random, independent pairs of observations $x_{11}, x_{21}, \cdots x_{1n}, x_{2n}$ are made on X_1 and X_2. The null hypothesis states that

$$F(x_1, x_2) \equiv f_1(x_1) \cdot f_2(x_2) \tag{6.1}$$

i.e., that X_1 and X_2 are independent The alternative hypotheses are that $F(x_1, x_2)$ does not satisfy (6.1).[1]

Let the set $x_{11}, x_{12}, x_{13}, \cdots, x_{1n}$ be arranged in order of ascending size, thus: $Z = z_1, z_2, z_3, \cdots, z_n$ where $z_1 < z_2 < \cdot < z_n$. The jth member of this sequence will be said to have the rank j. In the same manner ranks are assigned to the x_{2j} $(j = 1, \cdots, n)$. (It is easy to see that, since $f_1(x_1)$ and $f_2(x_2)$ are continuous, the probability that $z_j = z_{j+1}$ is 0 etc) In the sequence Z the element z_j $(j = 1, \cdot \cdot, n)$ is replaced by the rank of its associated observation on X_2. We obtain a permutation of the integers $1, 2, \cdot \ , n$ which we denote by R. If in the procedure for obtaining R, we had reversed the roles of the x_{1j} and x_{2j}, we would have obtained the permutation R' It is easy to see that any statistic, say M'', used to test the null hypothesis, must be a function only of R, with the added proviso that $M''(R) = M''(R')$ (The rank correlation coefficient is such a statistic) Under the null hypothesis all the R have the same probability $\left(=\frac{1}{n!}\right)$.

The procedure of applying the likelihood ratio principle to this problem would then be as follows. Ω is the totality of all bivariate distribution functions $H(x_1, x_2)$ which are continuous in both variables jointly The respective marginal distributions corresponding to $H(x_1, x_2)$ will be denotedby $h_1(x_1)$ and $h_2(x_2)$. ω is a subset of Ω which consists of all $H(x_1, x_2)$ for which $H(x_1, x_2) \equiv h_1(x_1) \cdot h_2(x_2)$. The sample space is the totality of all sequences R. The null hypothesis states that $F(x_1, x_2)$ is a member of ω. The admissible alternatives are that $F(x_1, x_2)$ is a member of Ω not in ω. The distribution of any function of R is the same for all members of ω. Thus the essential requirement for the applicability of the likelihood ratio criterion is fulfilled. All sequences R have the same probability for all members of ω; hence the numerator of the likelihood ratio is a func-

[1] It is easy to see that the independence or dependence of two stochastic variables is not a property which will remain invariant under a topologic transformation of the plane into itself. We therefore require of the statistic only that it be invariant under topologic transformation of *each* variable into itself, separately.

tion only of n which may therefore be ignored. We may then define $M'(R)$, a monotonic function of the likelihood ratio as the denominator of the likelihood ratio, thus:

$$M'(R) = \max_{\Omega} P\{R; H(x_1, x_2)\} \tag{6.2}$$

where $P\{R, H(x_1, x_2)\}$ is the probability of R when $H(x_1, x_2)$ is the joint distribution function of X_1 and X_2. The critical values of $M'(R)$ are the large values.

We now propose an approximation to $M'(R)$ which we shall call $M(R)$. We do this by describing a distribution function $H^*(x_1, x_2)$ for each R which seems a plausible approximation to a maximizing distribution function. It may be derived from certain assumptions about the nature of the maximizing distribution function which we omit. The remarks made in the preceding section about the character of the approximation apply here as well. As before we specify only the form of the function and leave certain parameters, finite in number, to be determined in accordance with (6.2). (If the construction of $H^*(x_1, x_2)$ should appear somewhat involved, this is due only to the analytic description. A sketch will show the essential simplicity of the situation.) We then have

$$M(R) = P\{R; H^*(x_1, x_2)\}.$$

Let $R = a_1, a_2, \cdots, a_n$ be a given permutation of the integers 1 to n. A sub-sequence $a_{(i+1)}, a_{(i+2)}, \cdots, a_{(i+l)}$ will be called a run of length l if the following conditions are fulfilled:

(6.3) The indices of the a's are consecutive,

(6.4) If l' is any integer such that $1 \leq l' < l$, then

$$| a_{(i+l')} - a_{(i+l'+1)} | = 1,$$

(6.5) if $i > 0$, $| a_i - a_{(i+1)} | > 1$,

(6.6) if $i + l < n$, $| a_{(i+l)} - a_{(i+l+1)} | > 1$.

The run will be called an ascending run or a descending run according as $a_{(i+1)} - a_{(i+2)} = -1$ or $+1$. A run of length 1 is of either type, at pleasure. For example, let

$$R = 5, 6, 1, 4, 3, 2.$$

The first run is 5, 6, the second 1, the last 4, 3, 2. 5, 6 is an ascending run of length two, 4, 3, 2 a descending run of length three, and 1 a run of length one.

$H^*(x_1, x_2)$ is a degenerate distribution function such that the relation between X_1 and X_2 is functional (this is a special case of stochastic relationship). That is to say, $X_2 = \varphi(X_1)$, where $\varphi(X_1)$ is a single-valued function defined for all the possible values of X_1, with a single-valued inverse $\varphi^{-1}(X_2)$ defined for all possible values of X_2. Hence $H^*(x_1, x_2)$ is completely specified when the function $X_2 = \varphi(X_1)$ and $h_1^*(x_1)$ the marginal distribution function of X_1, are given ($h_1^*(x_1)$ must of course be continuous).

Consider a system of intervals on the line $-\infty < x_1 < +\infty$ of which $(i - 1, i)$

is the ith, $i = 1, 2, \cdots, n$ and a similar system on the line $-\infty < x_2 < +\infty$. (Actually, as in the previous section, neither the length of the intervals nor their location is material The intervals need merely be disjunct and in a certain order. We are using these particular intervals to simplify the notation) Let l_1 be the length of the first run. a_1 is its first element. Then let

$$p_1 = P\{0 \leq X_1 \leq l_1 ; h_1^*(x_1)\}$$

be one of the as yet undetermined parameters We now partly define $h_1^*(x_1)$ as follows:

$$\begin{aligned} & h_1^*(x_1) = 0, & & x_1 \leq 0 \\ (6.7) \qquad & h_1^*(x_1) = 1, & & x_1 \geq n \\ & h_1^*(l_1) = p_1. & & \end{aligned}$$

Within the interval $(0, l_1)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6 7) We partly define $\varphi(X_1)$ as follows·
If the first run is ascending, let

$$\varphi(0) = a_1 - 1 \tag{6.8}$$

$$\varphi(x_1) = a_1 - 1 + x_1, \qquad 0 < x_1 \leq l_1. \tag{6.9}$$

If the first run is descending, let

$$\varphi(0) = a_1 \tag{6.10}$$

$$\varphi(x_1) = a_1 - x_1, \qquad 0 < x_1 \leq l_1. \tag{6.11}$$

We proceed in this manner through all the runs of R Let l_i be the length of the ith run. Let $\lambda_j = \sum_{i<j} l_i$. The first element of the jth run is $a_{(\lambda_j+1)}$. Let

$$p_j = P\{\lambda_j < X_1 \leq \lambda_j + l_j ; h_1^*(x_1)\},$$

be another of the as yet undetermined parameters. We then define $h_1^*(x_1)$ as follows:

$$h_1^*(\lambda_j) = \sum_{i<j} p_i \tag{6.12}$$

$$h_1^*(\lambda_j + l_j) = \sum_{i \leq j} p_i. \tag{6.13}$$

Within the interval $(\lambda_j, \lambda_j + l_j)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6.12) and (6 13). We define $\varphi(X_1)$ as follows:
If the jth run is ascending, let

$$\varphi(x_1) = a_{(\lambda_j+1)} - 1 + x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j). \tag{6.14}$$

If the jth run is descending, let

$$\varphi(x_1) = a_{(\lambda_j+1)} - x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j). \tag{6.15}$$

If $l_j = 1$, the run may be considered ascending or descending at pleasure.

In order to obtain R, it is necessary that all the elements of a run shall fall into its corresponding interval. Then it is easy to see that by the multinomial theorem

$$P\{R; H^*(x_1, x_2)\} = n! \prod_i (l_i!)^{-1} p_i^{l_i}. \tag{6.16}$$

The right member of (6.16) is to be maximized with respect to the p_i subject to the constraint

$$\Sigma p_i = 1. \tag{6.17}$$

It is easy to verify that the maximum occurs when

$$p_i = \frac{l_i}{n}. \tag{6.18}$$

$M(R)$ is the right member of (6.16) after the results (6.18) have been inserted. It is convenient to remove all factors which are functions only of n and to take the logarithm of the resulting expression. Then with a slight change of notation we may say that

$$M(R) = \sum_i \bar{l}_i \tag{6.19}$$

where

$$\bar{l}_i = \log\left(\frac{l_i^{l_i}}{l_i!}\right). \tag{6.20}$$

The critical values of $M(R)$ are the large values. One may verify without much difficulty that $M(R) = M(R')$, i.e., that the statistic is symmetric with respect to X_1 and X_2 as indeed it should be.

This result is immediately extensible to the problem of testing whether k stochastic variables $X_1, \cdots, X_k$ are independent. We shall not go into the details, which are similar to those described above, and content ourselves with giving the definition of a run for the case $k = 3$. After the observations on X_1 have been arranged in ascending order, we obtain two sequences R_2 and R_3, the associated ranks of the observations on X_2 and X_3. Let $R_2 = b_1, b_2, \cdots, b_n$ and $R_3 = b_1', b_2', \cdots, b_n'$. The ascending sequence of consecutive integers $(i + 1)$, $(i + 2)$, $\cdots$, $(i + l)$ determines a run of length l if the sequences $b_{(i+1)}, b_{(i+2)}, \cdots, b_{(i+l)}$ and $b'_{(i+1)}, b'_{(i+2)}, \cdots, b'_{(i+l)}$ *both* satisfy (6.4), and if at least one of the sequences satisfies (6.5), and at least one, but not necessarily the same one, satisfies (6.6). The adjectives ascending and descending apply to each sequence separately.

Let r_j be the number of runs of length j in R. Then

$$M(R) = \sum_j j r_j. \tag{6.21}$$

Most of the remarks made in Section 5 about the small sample distribution of $T(V)$ are also applicable to the distribution of $M(R)$. More will be said in the

next section about the distribution of $M(R)$ which involves the solution of a combinatorial problem not discussed in the literature.

7. On the distribution of $W(R)$. While most of the remarks made about the small sample distribution of $T(V)$ apply to the question of the distribution of $M(R)$ in small samples, the situation with respect to the distribution of $M(R)$ in samples of medium size and large size is very different and, in certain respects, is more favorable for practical application than is the case with $T(V)$ It would be reasonable to expect, for example, in view of Section 3 and of the structure of the statistic $M(R)$ that the asymptotic distribution of $M(R)$ should be normal. Surprisingly enough, this is not the case It is not even continuous In order to clarify the situation, we begin with a few necessary ideas and definitions.

Let the stochastic variable $W(R)$ be defined as the total number, in R, of runs of the sense of Section 6. We shall be interested in the distribution of $W(R)$. The number n of the pairs of observations on X_1 and X_2 (we consider the case of two variates) will be assumed arbitrary but fixed throughout the discussion and will not be exhibited Let $N(k)$ be the number of sequences R (of the integers 1 to n) which contain exactly k runs.

Consider, for example, for the case $n = 6$, the sequence 2 3 4 6 5 1. We shall say that this sequence contains the "contacts" (2, 3), (3, 4), (6, 5) In general, a contact is defined as the juxtaposition, in the sequence R, of consecutive numbers, whether in ascending or descending order. If k is the number of runs and l the number of contacts in a sequence R, then obviously

$$k + l = n. \tag{7 1}$$

Let R_0 be the sequence $1, 2, \cdots, n$ of the first n integers in ascending order. The $n - 1$ contacts of this sequence may themselves be arranged in a sequence R^* of contacts, thus·

$$(1, 2), (2, 3), \cdots, (n - 1, n).$$

Suppose l of the contacts which constitute the sequence R^* are selected in some manner to form the set O. The remaining $n - 1 - l$ contacts form the complementary set O'. After this selection the sequence R^* may be considered a sequence of the type of the sequences V of Section 5 with the members of O playing the role of the elements 0 and the members of O' playing the role of the elements 1 When R^* is considered in this manner we will write it as $R^*(O)$. The definition of a run of Section 5 as applied to sequences V is now applicable to $R^*(O)$. We will call any such run of the members of O or of O' a group.

We wish first to answer the following question· In how many ways can the set O be selected from among the elements of R^* so that it will contain l members arranged in $R^*(O)$ in i groups? If, for a given O, i' be the number of groups into which O' is divided in $R^*(O)$, it is clear that $i - i'$ can equal only -1, 0, or $+1$. Hence only four situations can arise, as follows.

a) $i' = i + 1$. The first group in $R^*(O)$ is therefore composed of elements of

O'. The number of ways in which l elements can be divided into i runs of the type of Section 2 is the coefficient of x^l in the purely formal expansion of

$$(x + x^2 + x^3 + \cdots)^i = \left(\frac{x}{1-x}\right)^i$$

and is therefore $\binom{l-1}{i-1}$. Similarly $n - 1 - l$ elements can be divided into $i' = i + 1$ runs in $\binom{n-l-2}{i}$ ways. Hence this situation will arise in $\binom{l-1}{i-1}\binom{n-l-2}{i}$ ways.

b) $i' = i - 1$. By a similar argument as above, this can occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-2}$ ways.

c) $i' = i$ and the first group is made up of elements from O. This will occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

d) $i' = i$ and the first group is made up of elements from O'. This will also occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

The set O which contains l elements arranged in i groups can therefore be selected in

$$\binom{l-1}{i-1}\left(\binom{n-l-2}{i} + \binom{n-l-2}{i-2} + 2\binom{n-l-2}{i-1}\right) \tag{7.2}$$

ways, and the quantity (7.2) is, by elementary combinatorics, equal to

$$\binom{l-1}{i-1}\binom{n-l}{i}. \tag{7.3}$$

Let any set O of l contacts divided into i groups be selected from R^*. Imagine that each contact in O sets up, in R_0, an unbreakable bond which links the two elements involved in the contact, but no contact in O' creates such a bond. Given these bonds set up by O, we seek the number of different sequences into which the n elements of R_0 can be permuted while respecting these bonds. Since there are l bonds, we can actually manipulate only $n - l$ entities, except that two elements linked by a bond may have their order reversed; for example, if O contains (1, 2), 1 may either precede or follow 2 and the bond would still be respected. However, if one contact in a group is reversed, the group as a whole must be reversed, else a bond would be broken. Hence the number of distinct sequences into which R_0 may be permuted while all the bonds set up by O are respected is $2^i(n - l)!$.

Let us refer to the sequences thus obtained as the family generated by O. All the sequences in a family are distinct. Now let O range over all sets of l

contacts selected from R^* The various families obtained will not be disjunct, but some will have sequences in common. In spite of this, we seek the total of the number of sequences in all the families. The total of the number of sequences in all the families generated by sets of l contacts divided into i groups is, by (7.3) and the result of the preceding paragraph,

$$2^i \cdot \binom{l-1}{i-1}\binom{n-l}{i}(n-l)!. \tag{7.4}$$

Sets of l contacts may consist of $1, 2, \cdots l$ groups, so that the total number of sequences in all the families generated by sets of l contacts is

$$A_l = \sum_{i=1}^{l} 2^i \binom{l-1}{i-1}\binom{n-l}{i}(n-l)! \tag{7.5}$$

where l may take the values $1, 2, \cdots, (n-1)$. The conventions on the combinatorial symbols will be:

$$\binom{a}{0} = 1, \qquad a \geq 0,$$

$$\binom{a}{b} = 0, \qquad a < b.$$

Define A_0 as

$$A_0 = n! \tag{7.6}$$

The following equation is trivial:

$$A_0 = \sum_{i=1}^{n} N(i). \tag{7.7}$$

We now consider all the families generated by sets O which contain exactly l contacts. As was said before, the total of the number of sequences in each is A_l Let $H(l)$ be the set of all the sequences in all these families, with each sequence in $H(l)$ counted as many times as the number of families in which it occurs. Every sequence in $H(l)$ has the l contacts of the set O which generated it, but after permuting R_0 other contacts may still exist. Hence every sequence in $H(l)$ has at least l contacts and therefore by (7.1), at most $n - l$ runs. Clearly, a sequence which has exactly l contacts occurs exactly once in $H(l)$, since it can appear only in the family generated by the set O of its l contacts and in no other family. A sequence which has exactly $(l + 1)$ contacts will appear exactly $\binom{l+1}{l}$ times in $H(l)$, for it will appear once in each family generated by a set O which consists of one of the $\binom{l+1}{l}$ selections of l contacts from among its $(l + 1)$ contacts, and in no other family Similarly each sequence which has exactly $(l + 2)$ contacts will appear in $H(l)$ $\binom{l+2}{l}$ times, and so forth. We therefore obtain, in view of (7.1),

$$A_l = \sum_{i=l}^{n-1} \binom{i}{l} N(n-i) \qquad (l = 1, 2, \cdots, (n-1)). \tag{7.8}$$

The system of n linear equations (7.7) and (7.8) completely determines the quantities $N(1), N(2), \cdots, N(n)$. The matrix of these equations has a determinant whose absolute value is 1, so that the quantities $N(1), N(2), \cdots, N(n)$ may readily be expressed in determinantal form. Furthermore the moments of $W(R)$ are readily found from these equations. Thus from (7.8) for $l = 1$ we find

$$E(W(R)) = \frac{n^2 - 2n + 2}{n} \sim n - 2 \tag{7.9}$$

and from (7.8) for $l = 2$ and $l = 1$ we find, after a little obvious manipulation,

$$\sigma^2(W(R)) = \frac{2n^3 - 8n^2 + 6n + 4}{n^3 - n^2} \sim 2. \tag{7.10}$$

Higher moments of $W(R)$ may be found in similar manner.

Since the limiting variance of $W(R)$ is 2 it follows that the asymptotic distribution is not continuous. For n of any size the bulk of the values are concentrated in a short interval ending at n. When $W(R) = n$, $M(R) = 0$, when $W(R) = n - 1$, $M(R) = \log 2$, and when $W(R) = n - 2$, $M(R) = \log 4\frac{1}{2}$ or $\log 4$. It is easy to see that for the values of $W(R)$ which differ very little from n there are only a small number of values of $M(R)$, whose asymptotic distribution is also discontinuous. The statistic $W(R)$ is therefore a good approximation to the statistic $M(R)$ for the purposes of tests of significance (for $M(R)$ the large values are the critical values and for $W(R)$ the small values are critical), and has a few additional practical advantages. It is even easier to compute than $M(R)$; the computation is best performed by counting contacts. Since the limiting variance is a small constant, it follows that many tests of significance can be performed simply by use of Tchebycheff's inequality. For example, suppose a given large sample contains 9 contacts, i.e., $n - 9$ runs (we say a "large" sample in order to use the simple limiting mean and variance; if desired or for a small sample these latter may be computed exactly by (7.9) and (7.10)). Then by Tchebycheff's inequality it follows that the probability of obtaining $n - 9$ or fewer runs is less than .041. Thus the presence of 9 contacts would be sufficient to render a sample of great size significant on a 5% level. For the few numbers of contacts about which doubt will exist as to whether or not they are critical values two procedures are possible. Either the equations (7.7) and (7.8) may be solved exactly for the doubtful values, or several higher moments may be found from (7.8) and the methods of Wald [14] can be applied to delimit the missing probabilities to any accuracy desired. By enumerating the few values of $M(R)$ which correspond to several of the largest values of $W(R)$ the distribution of $M(R)$ may be computed sufficiently to serve the purposes of tests of significance

REFERENCES

[1] MAURICE FRECHET, *Généralités sur les Probabilités, Variables Aleatoires*, Paris (1937)
[2] A. WALD and J WOLFOWITZ, *Annals of Math. Stat*, Vol 11 (1940), p. 147.
[3] A. M MOOD, *Annals of Math. Stat*, Vol 11 (1940), p. 367.
[4] J. NEYMAN, *Phil Trans. Roy. Soc London*, Vol 231 (1937), pp. 333–380
[5] A. WALD and J. WOLFOWITZ, *Annals of Math Stat*, Vol 10 (June, 1939), p 105
[6] H. HOTELLING and M PABST, *Annals of Math Stat*, Vol. 7 (1936), p. 29.
[7] E. G. OLDS, *Annals of Math Stat*, Vol 9 (1938), p. 133
[8] W R THOMPSON, *Annals of Math Stat.*, Vol. 9 (1938), p 281.
[9] W. J DIXON, *Annals of Math. Stat*, Vol. 11 (June, 1940), p 199.
[10] MILTON FRIEDMAN, *Jour Amer Stat Assoc*, Vol 32 (1937), p 675.
[11] J. NEYMAN and E. PEARSON, *Trans Royal Soc.*, A., Vol 231 (1933), p 295
[12] A. WALD, *Bull. Amer Math. Soc*, Vol 46 (1940), p. 235.
[13] A. WALD, *Bull Amer. Math. Soc*, Vol 47 (1941), p 396.
[14] A. WALD, *Trans. Amer. Math. Soc.*, Vol. 46 (1939), p 280.
[15] HARALD CRAMER, *Random Variables and Probability Distributions*, Cambridge (1937)
[16] F. HAUSDORFF, *Mengenlehre* (Second Edition), Berlin and Leipzig, 1927.

ON THE THEORY OF TESTING COMPOSITE HYPOTHESES WITH ONE CONSTRAINT

By Henry Scheffé

Princeton University

1. Introduction. Our purpose is to extend some of the Neyman-Pearson theory of testing hypotheses to cover certain cases of frequent interest which are complicated by the presence of nuisance parameters. Our results give methods of finding critical regions of types B and B_1. Type B regions were defined by Neyman [1] for the case of one nuisance parameter. Type B_1 regions are the natural generalization of the type A_1 regions of Neyman and Pearson [5] to permit the occurrence of nuisance parameters. The reader familiar with the work of these authors will recognize most of the notation and some of the methods.

We consider a joint distribution of n random variables $x_1, x_2, \cdots, x_n$, depending on l parameters $\theta_1, \theta_2, \cdots, \theta_l$, $l \leqq n$. The functional form of the distribution is given. The random variables may be regarded as the coordinates of a point E in an n-dimensional sample space W, the parameters, as the coordinates of a point Θ in an l-dimensional space Ω of admissible parameter values. Ω, unlike W, in general will not be a complete Euclidian space. Let ω denote the subspace of Ω defined by $\theta_1 = \theta_1^0$. The hypothesis we consider is

$$H_0 : \Theta \epsilon \omega.$$

Neyman and Pearson [4] call H_0 a hypothesis with $l - 1$ degrees of freedom; for our present purpose we shift the emphasis by saying it has one constraint.

It is clear that whenever we test whether a parameter has a given value, and other parameters occur in the distribution, we are testing a hypothesis with one constraint. Hypotheses of the type $\theta_1 = \theta_2$, in which we do not specify the common value of θ_1 and θ_2, nor the values of any other parameters, may always be transformed to H_0 by choosing new parameters. In general, the hypothesis that the parameter point Θ lies on some hypersurface in Ω, $g(\theta_1, \theta_2, \cdots, \theta_l) = g_0$, may be transformed to H_0 if the function g satisfies certain conditions,—say, g is continuous and monotone-increasing in one of the θ's for all Θ in Ω. Another circumstance lending importance to the theory of testing hypotheses with one constraint is its connection with the theory of confidence intervals, which we shall point out below.

The path which led Neyman to critical regions of type B is the following: Every Borel-measurable region w of sample space determines a test of H_0, which consists of rejecting H_0 if and only if E falls in w. In deciding which is a most efficient test, one may limit the competition to similar[1] regions, if such exist. Because of the general non-existence [2, p. 372] of uniformly most

[1] Defined by condition (a) of definition 1.

powerful tests, one is led to consider common best critical regions [4] if he is interested only in alternatives $\theta_1 < \theta_1^0$ (or $\theta_1 > \theta_1^0$), or else regions giving an unbiased test [1, p. 251]. Narrowing the competition further to the latter class of regions, one is led to regions of type B if he seeks tests which are most powerful for θ_1 very near to θ_1^0, and to type B_1 regions if he is not content with this. These types of regions are defined in section 2

We may now state the relationship of hypotheses with one constraint to the theory of confidence intervals [2]. To find confidence intervals for θ_1, we must first find similar regions $w(\theta_1^0)$ for testing H_0 If with every admissible θ_1 we can associate a $w(\theta_1)$, then confidence regions for θ_1 are determined, and if these be intervals, they are confidence intervals. Every class of similar regions mentioned above is intimately related to a category of confidence intervals. In particular, to find Neyman's short unbiased confidence intervals we must first solve the problem of type B regions. Likewise, if we define shortest unbiased confidence intervals in the obvious way along the lines laid down by Neyman, their discovery rests on the solution of the problem of type B_1 regions.

While the assumptions of section 3, especially 3^0, are unpleasantly restrictive—they are obviously tailored to fit the proof rather than the problem—they are nevertheless satisfied in many sampling problems associated with normal distributions. An application of the theorems of section 4 will be given in another paper *On the ratio of the variances of two normal populations*. The present theory was needed to round out that paper and was originally planned as a section thereof. However, it seems desirable for the convenience of other workers who might have use for the theory not to bury it under the preceding title.

Section 5 consists of an appendix on the moment problem raised by assumption 5^0.

2. Definitions. The symbols w, w_0, w_1 will always be understood to denote Borel-measurable regions in W. We shall symbolize $\partial^i Pr\{E \epsilon w \mid \Theta\}/\partial\theta_1^i$ for $i = 0, 1, 2$ by $P(w \mid \Theta)$, $P'(w \mid \Theta)$, $P''(w \mid \Theta)$, respectively. Since θ_1 plays a distinguished rôle, it will often be convenient to write $\Theta = (\theta_1, \vartheta)$, where the nuisance parameters are denoted by $\vartheta = (\theta_2, \theta_3, \cdots, \theta_l)$.

Definition 1: *w_0 is said to be a type B region for testing H_0 if for all Θ in ω*

(a) $P(w_0 \mid \theta_1^0, \vartheta) = \alpha$, *where α is independent of ϑ,*

(b) $P'(w_0 \mid \theta_1^0, \vartheta)$, $P''(w_0 \mid \theta_1^0, \vartheta)$ *exist,*

(c) $P'(w_0 \mid \theta_1^0, \vartheta) = 0$,

(d) $P''(w_0 \mid \theta_1^0, \vartheta) \geqq P''(w_1 \mid \theta_1^0, \vartheta)$ *for all w_1 satisfying* (a), (b), (c)

Definition 2: *w_0 is said to be of type B_1 if the conditions* (a), (b′), (c), (d′) *are satisfied. The conditions* (a), (c) *are given in definition* 1, *the other two are*

(b′) $P'(w_0 \mid \theta_1, \vartheta)$ *is continuous in θ_1 at $\theta_1 = \theta_1^0$ for all Θ in ω,*

(d′) $P(w_0^* \mid \theta_1, \vartheta) \geqq P(w_1 \mid \theta_1, \vartheta)$ *for all w_1 satisfying* (a), (b′), (c), *and all Θ in Ω.*

3. Assumptions. $p(x_1, x_2, \cdots, x_n \mid \Theta)$ will be a generic notation for the p.d.f. (probability density function) of random variables $x_1, x_2, \cdots, x_n$ whose distribution depends on Θ. The numbering of the following assumptions follows that of Neyman elsewhere [1].

1°. (a) There exists a p.d.f. $p(E \mid \Theta)$ such that for any w, and any $\Theta \in \Omega$,

$$P(w \mid \Theta) = \int_w p(E \mid \Theta)\, dW \tag{1}$$

where dW denotes the volume element $dx_1 dx_2 \cdots dx_n$.

(b) The region W_+ in W defined by $p(E \mid \Theta) > 0$ is independent of Θ for $\Theta \in \omega$.

(c) The connectivity of ω is such that it is possible to pass from any point Θ' in ω to any other point Θ'' in ω by a path lying entirely in ω and consisting of a finite number of segments on each of which all but one of $\theta_2, \theta_3, \cdots, \theta_l$ are constant.

2°. For all $E \in W_+$ and $\Theta \in \omega$, $p(E \mid \Theta)$ is differentiable twice with respect to θ_1 and indefinitely with respect to $\theta_2, \theta_3, \cdots, \theta_l$. For any w, and any $\Theta \in \omega$, the corresponding derivatives of $P(w \mid \Theta)$ exist and may be obtained by differentiating under the integral sign in (1).

We now define

$$\phi_i = \partial \log p(E \mid \Theta)/\partial\theta_i, \qquad \phi_{ij} = \partial\phi_i/\partial\theta_j, \qquad i, j = 1, 2, \cdots, l.$$

3°. For all $E \in W_+$ and $\Theta \in \omega$, $\phi_i = \phi_i(E, \Theta)$ is continuous in E, $i = 1, 2, \cdots, l$, and

$$\phi_{ij} = A_{ij} + \sum_{k=2}^{l} B_{ijk}\phi_k, \qquad i, j = 2, 3, \cdots, l, \tag{2}$$

$$\phi_{i1} = A_{i1} + \sum_{k=1}^{l} B_{i1k}\phi_k, \qquad i = 1, 2, \cdots, l, \tag{3}$$

where $A_{ij} = A_{ij}(\theta_1^0, \vartheta)$, $B_{ijk} = B_{ijk}(\theta_1^0, \vartheta)$ are continuous in each of $\theta_2, \theta_3, \cdots, \theta_l$.

4°. The matrix $(\partial\phi_i/\partial x_j)$, $i = 1, 2, \cdots, l$; $j = 1, 2, \cdots, n$, contains an $l \times l$ minor which is non-singular[2] for all $E \in W_+$ and $\Theta \in \omega$, and whose elements are continuous in E.

Write $\Phi = (\phi_2, \phi_3, \cdots, \phi_l)$, and denote by $p(\phi_1, \Phi \mid w, \Theta)$ the p.d.f. of (ϕ_1, Φ) calculated under the assumption that $E \in w$, i.e., that the p.d.f. of E is $p(E \mid \Theta)/P(w \mid \Theta)$ for $E \in w$ and zero for $E \in W - w$. Define

[2] If for each $\Theta \in \omega$, 4° is violated on an exceptional set $U(\Theta)$ for which $P(U(\Theta) \mid \Theta) = 0$, the theorems 1 and 2 may still be valid. What is essential is the existence of the p.d.f. $p(\phi_1, \phi_2, \cdots, \phi_l \mid \Theta)$ for all $\Theta \in \omega$. On reconsidering the theorems and their proofs, the reader will see that if the set $U(\Theta)$ is deleted from W_+, then 1°(b) may be violated, but not seriously, and no essential changes are necessary. The addition of the necessary qualifying clauses to our statements, regarding sets of probability zero, would encumber the developments.

$$Q_s(\Phi \mid w, \Theta) = \int_{-\infty}^{+\infty} \phi_1^s p(\phi_1, \Phi \mid w, \Theta) \, d\phi_1 \tag{4}$$

Let w_1 be any region satisfying condition (a) of definition 1.

5^0. We assume, for each $\Theta \epsilon \omega$, that if the moments[3] of $Q_s(\Phi \mid w_1, \Theta)$ and $Q_s(\Phi \mid W, \Theta)$ are the same then these functions are equal for almost all Φ

(a) for $s = 0$,

(b) for $s = 1$.

Note that Q_0 is p.d.f., Q_1 is not.

4. Theorems. A result of Neyman's [1] for $l = 2$ is generalized in the following[4]

THEOREM 1. *Under the assumptions 1^0 to 5^0, consider the existence of functions $k_i(\Phi, \theta_1^0, \vartheta)$, $i = 1, 2$, such that $k_1 < k_2$ and*

$$\int_{k_1(\Phi,\theta_1^0,\vartheta)}^{k_2(\Phi,\theta_1^0,\vartheta)} \phi_1^s p(\phi_1, \Phi \mid \theta_1^0, \vartheta) \, d\phi_1 = (1 - \alpha) \int_{-\infty}^{+\infty} \phi_1^s p(\phi_1, \Phi \mid \theta_1^0, \vartheta) \, d\phi_1, \qquad s = 0, 1, \tag{5}$$

for all $\Phi = (\phi_2, \phi_3, \cdots, \phi_l)$. If such functions exist for some $\Theta = \Theta' \epsilon \omega$, they exist for all $\Theta \epsilon \omega$. Then the region w_0 in W defined by

$$\phi_1(E, \theta_1^0, \vartheta) < k_1(\Phi, \theta_1^0, \vartheta) \quad \text{and} \quad \phi_1(E, \theta_1^0, \vartheta) > k_2(\Phi, \theta_1^0, \vartheta) \tag{6}$$

is independent of ϑ, and is a region of type B for testing the hypothesis H_0.

Since throughout the proof $\Theta = (\theta_1^0, \vartheta)$, we shall write Θ in place of these symbols to simplify the printing. It is to be understood that *every statement in the proof involving the symbol Θ is asserted for all Θ in ω.*

We suppose first that a type B region w_0 exists in W_+. Then from (a), (c) of definition 1 and assumptions 1^0(a) and 2^0,

$$\int_{w_0} p(E \mid \Theta) \, dW = \alpha, \tag{7}$$

$$\int_{w_0} \phi_1 p(E \mid \Theta) \, dW = 0. \tag{8}$$

Since the value of the integral (7) is independent of ϑ, all its derivatives with respect to $\theta_2, \theta_3, \cdots, \theta_l$ must vanish. This leads [3, pp. 50, 51. Insert k_t before ϕ_t^{-1} in (15)] to

---

[3] By this term we include "product moments."

[4] When I communicated this theorem to Professor Neyman, he informed me it was among the results of a thesis by R. Satô, *Contributions to the theory of testing statistical composite hypotheses*, University of London, 1937, and he kindly sent me a copy of the MS. I decided nevertheless to publish my version of theorem and proof, since for the reasons indicated in section 1 this theory should be available in the literature.

$$(9) \qquad \alpha^{-1} \int_{w_0} \prod_{i=2}^{l} \phi_i^{k_i} p(E \mid \Theta)\, dW = M(k_2, k_3, \cdots, k_l \mid \Theta), \qquad k_i = 0, 1, 2, \cdots,$$

where M is independent of w_0, and thus has the value obtained from (9) by putting $w_0 = W$ and $\alpha = 1$. In particular,

$$(10) \qquad \alpha^{-1} \int_{w_0} \phi_i p(E \mid \Theta)\, dW = 0, \qquad i = 2, 3, \cdots, l.$$

The necessary condition (9) for (7) is also sufficient. Denoting by $\mathcal{E}(f \mid w, \Theta)$ the expected value of a function $f(E, \Theta)$ calculated under the assumption that $E \in w$, equation (9) may be written

$$(11) \qquad \mathcal{E}\left(\prod_{i=2}^{l} \phi_i^{k_i} \mid w_0, \Theta\right) = \mathcal{E}\left(\prod_{i=2}^{l} \phi_i^{k_i} \mid W, \Theta\right).$$

From assumption 5^0(a) it then follows that

$$(12) \qquad Q_0(\Phi \mid w_0, \Theta) = Q_0(\Phi \mid W, \Theta)$$

for almost all Φ. Conversely, (12) implies (11).

In a similar manner we get from (8) with the aid of (9),

$$(13) \qquad \mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid w_0, \Theta\right) = \mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid W, \Theta\right).$$

We calculate the moments of the function $Q_1(\Phi \mid w, \Theta)$ to be

$$\mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid w, \Theta\right),$$

and hence because of 5^0(b), (13) implies

$$(14) \qquad Q_1(\Phi \mid w_0, \Theta) = Q_1(\Phi \mid W, \Theta)$$

almost everywhere in the Φ-space. The pair of conditions (12), (14) are equivalent to the pair (7), (8).

In order that w_0 be a type B region, it is necessary and sufficient that it satisfy (12) and (14) and that

$$P''(w_0 \mid \Theta) \geqq P''(w_1 \mid \Theta)$$

for all w_1 satisfying (12) and (14). The inequality may be transformed with the help of 1^0(a), 2^0, (3), (7), (8), and (10) to

$$\int_{w_0} \phi_1^2 p(E \mid \Theta)\, dW \geqq \int_{w_1} \phi_1^2 p(E \mid \Theta)\, dW,$$

which is equivalent to

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_1^2 p(\phi_1, \Phi \mid w_0, \Theta)\, d\phi_1\, d\phi_2 \cdots d\phi_l$$
$$\geqq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_1^2 p(\phi_1, \Phi \mid w_1, \Theta)\, d\phi_1\, d\phi_2 \cdots d\phi_l.$$

Sufficient for this is

$$Q_2(\Phi \mid w_0, \Theta) \geqq Q_2(\Phi \mid w_1, \Theta). \tag{15}$$

We note the functions in (12), (14), and (15) are all of the form (4) with $s = 0, 1, 2$, and propose to transform these to integrals over certain portions of the sample space W. First, we write (4) in the form

$$Q_0(\Phi \mid w, \Theta)\int_{-\infty}^{+\infty} \phi_1^s p(\phi_1 \mid \Phi, w, \Theta)\, d\phi_1 = Q_0(\Phi \mid w, \Theta)\mathcal{E}(\phi_1^s \mid \Phi, w, \Theta) \tag{16}$$

Next, we consider "surfaces" $S(\Phi, \Theta)$ in W_+, constructed as follows For any fixed Θ let $D(\Theta)$ be the $l - 1$ dimensional domain of values of $\phi_i(E, \Theta)$, $i = 2, 3, \cdots, l$, for $E \epsilon W_+$. A "surface" $S(\Phi, \Theta)$ is the locus of points E for which

$$\phi_i(E, \Theta) = \phi_i', \text{ a constant}, \qquad i = 2, 3, \cdots, l, \tag{17}$$

the set of constants being in $D(\Theta)$. Over every "surface" we now define a density ρ. Without loss of generality, and to simplify the notation, we shall assume that the non-singular minor postulated in 4^0 contains the minor $(\partial\phi_i/\partial x_j)$, $i = 2, 3, \cdots, l$; $j = 1, 2, \cdots, l - 1$, and denote by $J(E, \Theta)$ its determinant. For E on $S(\Phi, \Theta)$ we define the density

$$\rho(E \mid \Theta) = p(E \mid \Theta)/ \mid J(E, \Theta) \mid, \tag{18}$$

and consider "surface" integrals

$$\int_{wS(\Phi,\Theta)} F_s(E, \Theta)\, dx_l\, dx_{l+1} \cdots dx_n, \tag{19}$$

where

$$F_s(E, \Theta) = \phi_1^s(E, \Theta)\rho(E \mid \Theta). \tag{20}$$

A "surface" integral (19) is to be distinguished from an ordinary multiple integral, in that the integrand is not merely a function of $x_l, x_{l+1}, \cdots, x_n$; there may be several points E on the surface with the same values for these coordinates, but different values for the integrand. The integral is to be thought of as follows: The part $wS(\Phi, \Theta)$ of the "surface" $S(\Phi, \Theta)$ is partitioned into pieces ΔS, on each a point E is chosen, and the value of the integrand at E is multiplied by the "area" of the projection (taken non-negative) of ΔS on the $x_l, x_{l+1}, \cdots, x_n$-space. The "surface" integral is the limit of the sum of such products as the norm of the partition approaches zero.

Denoting the integral (19) by $I(s)$ for the moment, we may calculate that for $\Phi \epsilon D(\Theta)$

$$I(s) = I(0)\mathcal{E}(\phi_1^s \mid \Phi, w, \Theta), \qquad I(0) = Q_0(\Phi \mid w, \Theta)P(w \mid \Theta),$$

and hence we see that the right member of (16) is equal to the integral (19) divided by $P(w \mid \Theta)$. The desired relationship between the ordinary integrals (4) and the "surface" integrals (19) is thus

$$Q_s(\Phi \mid w, \Theta) = \int_{wS(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j / P(w \mid \Theta). \tag{21}$$

The conditions (12), (14), (15) may now be written

$$\int_{w_0 S(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j = \alpha \int_{S(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j, \qquad s = 0, 1, \tag{22}$$

$$\int_{w_0 S(\Phi,\Theta)} F_2(E, \Theta) \prod_{j=l}^{n} dx_j \geqq \int_{w_1 S(\Phi,\Theta)} F_2(E, \Theta) \prod_{j=l}^{n} dx_j, \tag{23}$$

if Φ is in the domain $D(\Theta)$, else they are satisfied trivially. w_0 will be a type B region if equations (22) are satisfied for almost all $\Phi \in D(\Theta)$, and if (23) is valid for all w_1 satisfying (22).

We now hold Θ fixed in ω and Φ fixed in $D(\Theta)$, so that $S(\Phi, \Theta)$ is fixed, and the right members of equations (22) have constant values. The proof [5, p. 11] of the lemma of Neyman and Pearson giving sufficient conditions that a region maximize an integral, subject to integral side-conditions, is easily seen to be valid for our "surface" integrals, and a sufficient condition that $w_0 S(\Phi, \Theta)$ have the desired property is then that it be defined by

$$\phi_1^2(E, \Theta) > a_0 + a_1 \phi_1(E, \Theta), \tag{24}$$

where a_0, a_1 are independent of E on $S(\Phi, \Theta)$, and are such that equations (22) are satisfied. Since Θ and Φ are fixed, we may permit a_i to be of the nature $a_i = a_i(\Phi, \Theta)$, $i = 1, 2$ Introducing functions $k_1 < k_2$, $k_i = k_i(\Phi, \Theta)$, and defining a_0, a_1 from

$$a_0 = -k_1 k_2, \qquad a_1 = k_1 + k_2,$$

the inequality (24) is satisfied if (6) is. Still holding Θ fixed, suppose that k_1, k_2 can be determined for all Φ (hence almost all Φ) in $D(\Theta)$ so that for the part $w_0 S(\Phi, \Theta)$ of $S(\Phi, \Theta)$, defined by (6), the equations (22) are satisfied. The parts $w_0 S(\Phi, \Theta)$ of "surfaces" then sweep out a "solid" $w_0(\Theta)$ in W_+, defined by (6). If we can similarly determine k_1 and k_2, and hence $w_0(\Theta)$, for every Θ in ω, and if furthermore $w_0(\Theta)$ is independent of Θ, then it is the type B region we seek.

The equations (22) have now served their main purpose, and we return to their equivalents, (12) and (14). For $w_0(\Theta)$ defined by (6)

$$p(\phi_1, \Phi \mid w_0, \Theta) = p(\phi_1, \Phi \mid W, \Theta)/\alpha \quad \text{if} \quad \phi_1 < k_1 \quad \text{or} \quad \phi_1 > k_2,$$

and vanishes otherwise, and hence equations (12) and (14) are equivalent to (5).

The remainder of the proof consists of deducing that k_1, k_2 exist, and that the associated region $w_0(\Theta)$ is independent of Θ, for all $\Theta \in \omega$, from the hypothesis of our theorem that k_1, k_2 exist for some $\Theta = \Theta'$. By 1^0(c), Θ' lies on a line segment L entirely in ω, on which all but one of the nuisance parameters, say θ_2, are constant Let us vary Θ over L. Then θ_3, θ_4, $\cdots$, θ_l remain fixed and θ_2 varies over an interval I. The equations (2) for $j = 2$ now become

ordinary differential equations in which the independent variable is θ_2, the dependent variables are $\phi_2, \phi_3, \cdots, \phi_l$, and $\theta_1^0, \theta_3, \cdots, \theta_l$ are parameters. A well known existence theorem assures us of the existence of particular solutions u_i and a non-singular (for all θ_2 in I) matrix (u_{ij}) of complementary solutions, $i, j = 2, 3, \cdots, l$, such that the general solution is

$$\phi_i = u_i + \sum_{j=2}^{l} u_{ij} c_j .$$

The u_i are determined by initial conditions for the system (2) with $j = 2$, and the u_{ij} by sets of initial conditions for the corresponding complementary system. Clearly, if these initial conditions are all chosen independent of E, then since the coefficients of the differential equations are all independent of E, the solutions u_i and u_{ij} enjoy the same property. On the other hand, the c_j are independent of θ_2. Hence

$$\phi_i(E, \theta_2) = u_i(\theta_2) + \sum_{j=2}^{l} u_{ij}(\theta_2)c_j(E), \qquad i = 2, 3, \cdots, l. \tag{25}$$

The dependence of the ϕ's, u's and c's on the parameters $\theta_1^0, \theta_3, \cdots, \theta_l$ has not been indicated, since these remain fixed throughout the present calculations.

Let $\mathfrak{D}$ be the $l - 1$ dimensional domain of the values of $c_j(E)$ for $E \epsilon W_+$, and C: $(c_2', c_3', \cdots, c_l')$ be a point in $\mathfrak{D}$, and denote by $S(C)$ the "surface" $c_j(E) = c_j'$. Denote the surface $S(\Phi, \Theta)$ defined in (17) by $S(\Phi, \theta_2)$, and the domain $D(\Theta)$ of Φ by $D(\theta_2)$. Then since $| u_{ij} | \neq 0$, therefore for every $\theta_2 \epsilon I$, every $S(C)$ with $C \epsilon \mathfrak{D}$ is identical with some $S(\Phi, \theta_2)$ with $\Phi \epsilon D(\theta_2)$, and vice versa. From this we conclude for later reference: (A) the functions $c_j(E)$ are constant on every $S(\Phi, \theta_2)$; (B) if θ_2', θ_2'' are any two values in I, then for every $\Phi = \Phi'' \epsilon D(\theta_2'')$ there exists a $\Phi' \epsilon D(\theta_2')$ such that $S(\Phi', \theta_2')$ is identical with $S(\Phi'', \theta_2'')$, and vice versa.

Now let us integrate with respect to θ_2 the equation

$$\partial \log p(E \mid \theta_2)/\partial\theta_2 = \phi_2 = u_2(\theta_2) + \sum_{j=2}^{l} u_{2j}(\theta_2)c_j(E).$$

$$\log p(E \mid \theta_2) = v(\theta_2) + \sum_{j=2}^{l} v_j(\theta_2)c_j(E) + f(E),$$

where $v(\theta_2)$, $v_j(\theta_2)$, $f(E)$, and all new undefined symbols in the sequel have obvious meanings. We get

$$p(E \mid \theta_2) = \bar{v}(\theta_2)\bar{f}(E) \exp\left[\sum_{j=2}^{l} v_j(\theta_2)c_j(E)\right]. \tag{26}$$

Next we differentiate the equations (25) with respect to x_k, and write the result in matrix form,

$$(\partial\phi_i/\partial x_k) = (u_{ij})(\partial c_j/\partial x_k), \qquad i, j = 2, 3, \cdots, l; k = 1, 2, \cdots, l - 1.$$

Taking determinants, we have

$$J(E, \theta_2) = J_1(\theta_2)J_2(E). \tag{27}$$

Finally, we shall need to know the nature of the dependence of ϕ_1 on θ_2 and E: From (3),

$$\partial\phi_1/\partial\theta_2 = A_{12}(\theta_2) + B_{121}(\theta_2)\phi_1 + \sum_{j=2}^{l} B_{12k}(\theta_2)\phi_k .$$

Substituting from (25), we get

$$\partial\phi_1/\partial\theta_2 = B_{121}(\theta_2)\phi_1 + A(\theta_2) + \sum_{j=2}^{l} B_j(\theta_2)c_j(E),$$

and integrating,

$$\phi_1(E, \theta_2) = B(\theta_2)\left[\int^{\theta_2} \frac{A(\xi) + \sum_{j=2}^{l} B_j(\xi)c_j(E)}{B(\xi)}\, d\xi + g(E)\right],$$

where

$$(28)\qquad B(\theta_2) = \exp\left[\int^{\theta_2} B_{121}(\eta)\, d\eta\right].$$

Thus

$$(29)\qquad \phi_1(E, \theta_2) = \bar{A}(\theta_2) + \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E) + B(\theta_2)g(E).$$

In equations (22) we now use the definitions (20), (18) for the integrands and then substitute (26), (27), (29). As a result we obtain the equality of

$$\int_{\omega_0 S(\Phi,\theta_2)} \frac{\left[\bar{A}(\theta_2) + \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E) + B(\theta_2)g(E)\right]^s \bar{v}(\theta_2)\bar{f}(E) \cdot\exp\left[\sum_{j=l}^{l} v_j(\theta_2)c_j(E)\right]}{|J_1(\theta_2)J_2(E)|} \prod_{j=l}^{n} dx_j$$

and α times the "surface" integral of the same integrand over $S(\Phi, \theta_2)$. Putting first $s = 0$ and then $s = 1$, and employing the previous conclusion (A), we find that the equations (22) are equivalent to

$$(30)\qquad \begin{aligned}&\int_{\omega_0 S(\Phi,\theta_2)} \{g^s(E)\bar{f}(E)/|J_2(E)|\} \prod_{j=l}^{n} dx_j\\ &\qquad = \alpha \int_{S(\Phi,\theta_2)} \{g^s(E)\bar{f}(E)/|J_2(E)|\} \prod_{j=l}^{n} dx_j, \qquad s = 0, 1.\end{aligned}$$

Again using the expression (29) for ϕ_1, and noting from (28) that $B(\theta_2) > 0$, we may write the inequality (6) in the form

$$(31)\qquad g(E) < \kappa_1(\Phi, \theta_2) \quad\text{and}\quad g(E) > \kappa_2(\Phi, \theta_2),$$

where

$$(32)\qquad \kappa_i(\Phi, \theta_2) = \left[k_i(\Phi, \theta_1^0, \vartheta) - \bar{A}(\theta_2) - \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E)\right] \Big/ B(\theta_2).$$

It follows from our hypothesis that for $\theta_2 = \theta_2'$ (the θ_2 coordinate of Θ') and any $\Phi \epsilon D(\theta_2')$, functions $\kappa_i(\Phi, \theta_2')$ exist such that for the part $w_0S(\Phi, \theta_2')$ of $S(\Phi, \theta_2')$, defined by (31), equations (30) are satisfied. The region $w_0(\Theta')$ is "swept out" by $w_0S(\Phi, \theta_2')$ as Φ ranges over $D(\theta_2')$. Now let Θ'' be any other $\Theta \epsilon L$, call its θ_2 coordinate θ_2'', let Φ'' be any $\Phi \epsilon D(\theta_2'')$, and consider the possibility of finding $\kappa_i(\Phi'', \theta_2'')$ such that on the part $w_0S(\Phi'', \theta_2'')$ of $S(\Phi'', \theta_2'')$, defined by (31), equations (30) are satisfied. From the conclusion (B), $S(\Phi'', \theta_2'')$ is identical with $S(\Phi', \theta_2')$ for a suitably chosen $\Phi' \epsilon D(\theta_2')$. Hence if we take $\kappa_i(\Phi'', \theta_2'') = \kappa_i(\Phi', \theta_2')$, then $w_0S(\Phi'', \theta_2'')$ becomes identical with $w_0S(\Phi', \theta_2')$ where equations (30) are already satisfied. Letting Φ'' range over $D(\theta_2'')$, every $w_0S(\Phi'', \theta_2'')$ thus determined becomes identical with some $w_0S(\Phi', \theta_2')$, and vice versa, by (B). Thus the region $w_0(\Theta'')$ "swept out" is identical with $w_0(\Theta')$. This process defines $\kappa_i(\Phi, \theta_2)$ for all $\theta_2 \epsilon I$ and $\Phi \epsilon D(\theta_2)$, and hence determines $k_i(\Phi, \theta_1^0, \vartheta)$ from (32). We now have functions $k_i(\Phi, \theta_1^0, \vartheta)$, $k_1 < k_2$, satisfying (5), and corresponding regions $w_0(\Theta)$ independent of Θ, for all $\Theta \epsilon L$. To conclude the proof, we use $1^0(c)$ to reach any point Θ in ω from Θ' by a path consisting of a finite number of segments like L on which only one of the nuisance parameters varies. The definitions of $k_i(\Phi, \theta_1^0, \vartheta)$ are continued along this path as above and the region $w_0(\Theta)$ is seen to be independent of Θ for all Θ in ω.

The following theorem may be regarded as a generalization of one by Neyman [6, p. 33] giving sufficient conditions that a type A region be also of type A_1:

THEOREM 2. *Suppose the assumption $1^0(b)$ holds for all $\Theta \epsilon \Omega$. Denote $\phi_i(E, \theta_1^0, \vartheta)$ by ϕ_i^0 and let $R(\vartheta)$ be the domain of values of $\phi_1^0, \phi_2^0, \cdots, \phi_l^0$ for $E \epsilon W_+$ and $\Theta \epsilon \omega$. Then a sufficient condition that a region w_0 of type B, found by application of theorem 1, be also of type B_1 is that for all $\Theta \epsilon \Omega$ and all $E \epsilon W_+$*

$$(33) \qquad p(E \mid \theta_1, \vartheta) = p(E \mid \theta_1^0, \vartheta) g(\phi_1^0, \phi_2^0, \cdots, \phi_l^0; \theta_1^0; \theta_1, \vartheta),$$

where $g(y_1, y_2, \cdots, y_l; \theta_1^0; \theta_1, \vartheta)$ is a function such that $\partial^2 g/\partial y_1^2 > 0$ for all $y_1, y_2, \cdots, y_l$ in $R(\vartheta)$ and $\Theta \epsilon \Omega - \omega$.

For the w_0 satisfying the sufficient conditions of theorem 1, the conditions (a), (b′), (c) of definition 2 are satisfied, and it remains only to verify the condition (d′). The regions w_1 admitted for comparison in (d), as well as w_0, must satisfy the equations (22) since these are equivalent to the conditions (a), (c). We recall that $\Theta = (\theta_1^0, \vartheta)$ in equations (22) and rewrite them in a notation better adapted to our present considerations:

$$(34) \qquad \int_{w_0S(\Phi^0,\theta_1^0,\vartheta)} [\phi_1^0]^s \{p(E \mid \theta_1^0, \vartheta)/\mid J(E, \theta_1^0, \vartheta) \mid \} \prod_{j=l}^{n} dx_j$$
$$= \alpha \int_{S(\Phi^0,\theta_1^0,\vartheta)} [\phi_1^0]^s \{p(E \mid \theta_1^0, \vartheta)/\mid J(E, \theta_1^0, \vartheta) \mid \} \prod_{j=l}^{n} dx_j, \quad s = 0, 1$$

where $\Phi^0 = (\phi_2^0, \phi_3^0, \cdots, \phi_l^0) \epsilon D(\theta_1^0, \vartheta)$.

To express the condition (d) in a convenient way, we now "shred" the regions w_0, w_1 of (d) for every θ_1 by means of the same "surfaces" we have been using

for $\theta_1 = \theta_1^0$: For any w in W_+, $\Theta \epsilon \Omega$, and $\Phi^0 \epsilon D(\theta_1^0, \vartheta)$ we define a "surface" integral

$$I(\Phi^0, w \mid \theta_1, \vartheta) = \int_{wS(\Phi^0, \theta_1^0, \vartheta)} \{p(E \mid \theta_1, \vartheta)/|J(E, \theta_1^0, \vartheta)|\} \prod_{j=l}^{n} dx_j.$$

Then

$$P(w \mid \theta_1, \vartheta) = \int \cdots \int_{D(\theta_1^0, \vartheta)} I(\Phi^0, w \mid \theta_1, \vartheta)\, d\phi_2^0\, d\phi_3^0 \cdots d\phi_l^0,$$

and a sufficient condition for (d) is

$$I(\Phi^0, w_0 \mid \theta_1, \vartheta) \geqq I(\Phi^0, w_1 \mid \theta_1, \vartheta) \tag{35}$$

for all $\Theta \epsilon \Omega$ and all $\Phi^0 \epsilon D(\theta_1^0, \vartheta)$.

Again applying the lemma of Neyman and Pearson to the integrands of the "surface" integrals in (34) and (35), we find that a sufficient condition that our region w_0 be of type B_1 is that there exist functions $b_i(\Phi^0, \theta_1^0, \theta_1, \vartheta)$, $i = 1, 2$, such that

$$p(E \mid \theta_1, \vartheta) > p(E \mid \theta_1^0, \vartheta)[b_0 + b_1\phi_1^0(E, \theta_1^0, \vartheta)]$$

if and only if $E \epsilon w_0$. Employing (33), we may replace this inequality by

$$g(\phi_1^0, \Phi^0; \theta_1^0; \theta_1, \vartheta) > b_0 + b_1\phi_1^0. \tag{36}$$

Define b_0, b_1 from

$$g(k_i, \Phi^0; \theta_1^0; \theta_1, \vartheta) = b_0 + b_1 k_i, \qquad i = 1, 2,$$

where $k_i = k_i(\Phi^0, \theta_1^0, \vartheta)$. Since $k_1 < k_2$, these equations have unique solutions b_0, b_1. Now hold Φ^0, θ_1, ϑ all fixed ($\theta_1 \neq \theta_1^0$) and consider the graphs of the members of (36) as functions of ϕ_1^0. From our definition of b_0, b_1, these graphs intersect at k_1, k_2. But by hypothesis, the graph of the left member is everywhere concave up, and hence for $k_1 < \phi_1^0 < k_2$, it lies below the linear graph of the right member, and for $\phi_1^0 < k_1$ and $\phi_1^0 > k_2$, it lies above. That is (36) is true if and only if $E \epsilon w_0$.

5. Appendix on the moment problem. Easily applied criteria [8] are available for the moment problem of assumption $5^0(a)$. The moment problem $5^0(b)$ is much more difficult, however, because the function to be determined by its moments is not of constant sign. Below we offer a proof that the solutions of both problems $5^0(a)$ and $5^0(b)$ are unique in the important case where $p(E \mid \Theta)$ is a multivariate normal p.d.f. and $\phi_1, \phi_2, \cdots, \phi_l$ are polynomials in $x_1, x_2, \cdots, x_n$ of degree $\leqq 2$ and not necessarily homogeneous. Since Θ is held fixed, we will not indicate dependence on Θ, nor will the dependence of various functions on s be indicated, since $s = 0$ or else 1 throughout.

Let w_1, w_2 be any two regions, $\alpha_j = P(w_j) \neq 0$, for which the moments of $Q_s(\Phi \mid w_1)$ and $Q_s(\Phi \mid w_2)$ are the same. To prove the equality (almost every-

where) of these two functions it suffices to prove that their Fourier transforms are identical [7, theorem 61]. Suppressing the customary multiple of $\sqrt{2\pi}$, the Fourier transform of $Q_s(\Phi \mid w_j)$ is

$$\Psi_j(\mathbf{t}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\mathbf{t}\cdot\Phi} Q_s(\Phi \mid w_j)\, d\phi_2 \cdots d\phi_l,$$

where $\mathbf{t}$ is the vector $(t_2, t_3, \cdots, t_l)$ and $\mathbf{t}\cdot\Phi = t_2\phi_2 + \cdots + t_l\phi_l$. From (4) we get

$$\begin{aligned}\Psi_j(\mathbf{t}) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\mathbf{t}\cdot\Phi} \phi_1^s p(\phi_1, \Phi \mid w_j)\, d\phi_1\, d\phi_2 \cdots d\phi_l \\ &= \mathcal{E}(e^{i\mathbf{t}\cdot\Phi} \phi_1^s \mid w_j) \\ &= \frac{1}{\alpha_j} \int_{w_j} e^{i\mathbf{t}\cdot\Phi} \phi_1^s(E) p(E)\, dW.\end{aligned}$$

A device of Cramér and Wold [8] for reducing the dimensionality of the problem now suggests itself. Let z be a scalar variable and consider $\psi_j(z \mid \mathbf{t}) = \Psi_j(z\mathbf{t})$ for fixed $\mathbf{t}$ as a function of z. Obviously if for every fixed $\mathbf{t}$, $\psi_1(z \mid \mathbf{t}) = \psi_2(z \mid \mathbf{t})$, then $\Psi_1(\mathbf{t}) = \Psi_2(\mathbf{t})$, and we are through. We propose to prove the former equality by showing first that ψ_j is an analytic function of z for all real z and secondly that the coefficients of the power series for ψ_1 and ψ_2 in powers of z are equal. Holding $\mathbf{t}$ fixed now, $\xi = \mathbf{t}\cdot\Phi$ is a polynomial of degree $\leqq 2$, and

$$\psi_j(z \mid \mathbf{t}) = \frac{1}{\alpha_j} \int_{w_j} e^{iz\xi} \phi_1^s p\, dW. \tag{37}$$

By our assumption of normality,

$$p = C \exp\left[-\sum_{\kappa,\nu=1}^{n} a_{\kappa\nu} y_\kappa y_\nu \right], \qquad y_\kappa = x_\kappa - \mu_\kappa,$$

where the matrix $(a_{\kappa\nu})$ is positive definite. To prove the analyticity of ψ_j for any real $z = z_0$, let $z = z_0 + \zeta$, and restrict ζ to real values. Substitute in (37)

$$e^{i\zeta\xi} = \sum_{q=0}^{m-1} \frac{(i\zeta\xi)^q}{q!} + \frac{(i\zeta\xi)^m}{m!} f_m(\zeta\xi),$$

where $|f_m(\zeta\xi)| \leqq 1$. Then

$$\psi_j(z_0 + \zeta \mid \mathbf{t}) = \sum_{q=0}^{m-1} \frac{(i\zeta)^q}{q!\,\alpha_j} \int_{w_j} e^{iz_0\xi} \xi^q \phi_1^s p\, dW + R_{jm}(z_0, \zeta),$$

where

$$R_{jm} = \frac{(i\zeta)^m}{m!\,\alpha_j} \int_{w_j} e^{iz_0\xi} f_m(\zeta\xi) \xi^m \phi_1^s p\, dW,$$

and all integrands are absolutely integrable over W. Let σ be the sphere of unit radius with center at $(\mu_1, \mu_2, \cdots, \mu_n)$ in W and write

$$R_{jm} = \frac{(i\zeta)^m}{m!\,\alpha_j}\left[\int_{w_j\sigma} + \int_{w_j - w_j\sigma}\right].$$

Call the two terms of the right member R'_{jm} and R''_{jm},

$$R_{jm} = R'_{jm} + R''_{jm}.$$

$$|R'_{jm}| \leqq \frac{|\zeta|^m}{m!\,\alpha_j}\int_{w_j\sigma} |\xi^m \phi_1^s|\, p\, dW.$$

Let $M = \max|\xi|$, $M_1 = \max|\phi_1^s|$, for $E \epsilon \sigma$. Then

$$|R'_{jm}| \leqq \frac{M_1|M\zeta|^m}{m!\,\alpha_j}\int_\sigma p\, dW \leqq M_1|M\zeta|^m/m!\,\alpha_j.$$

Hence $R'_{jm} \rightarrow 0$ for all real ζ as $m \rightarrow \infty$.

$$|R''_{jm}| \leqq \frac{|\zeta|^m}{m!\,\alpha_j}\int_{W-\sigma} |\xi^m \phi_1^s|\, p\, dW.$$

Let $r = \left(\sum_{\kappa=1}^{n} y_\kappa^2\right)^{\frac{1}{2}}$, and M_2, M_3 be the sums of the absolute values of the coefficients of the polynomials ϕ_1^s, ξ, respectively, when expanded in powers of y_κ. Then for $E \epsilon W - \sigma$, $|\phi_1^s| \leqq M_2 r^2$, $|\xi| \leqq M_3 r^2$, $p \leqq C \exp(-\lambda r^2)$, where $\lambda > 0$ is the smallest characteristic root of $(a_{\kappa\nu})$. Hence

$$|R''_{jm}| \leqq \frac{CM_2|M_3\zeta|^m}{m!\,\alpha_j}\int_{W-\sigma} r^{2m+2} e^{-\lambda r^2}\, dW.$$

Integrating over spherical shells concentric with σ, $dW = M_4 r^{n-1}\, dr$, and

$$|R''_{jm}| \leqq \frac{CM_2M_4|M_3\zeta|^m}{m!\,\alpha_j}\int_1^\infty r^{2m+n+1} e^{-\lambda r^2}\, dr \leqq \frac{CM_2M_4|M_3\zeta|^m}{m!\,\alpha_j}\int_0^\infty.$$

If we evaluate the last integral in terms of a Gamma function and employ Stirling's formula we easily find that for $M_3|\zeta| < \lambda$, $R''_{jm} \rightarrow 0$. The convergence of R_{jm} to zero for *real* ζ, $|\zeta| < \lambda/M_3$, is sufficient to insure the analyticity of ψ_j.

Now let $z_0 = 0$. Then the coefficient of ζ^q in the power series for ψ_j is

$$\frac{i^q}{q!\,\alpha_j}\int_{w_j} (t_2\phi_2 + \cdots + t_l\phi_l)^q \phi_1^s p\, dW,$$

a linear combination (the same for $j = 1, 2$) of the q-th order moments of $Q_s(\Phi \mid w_j)$, and hence corresponding coefficients for ψ_1 and ψ_2 are equal.

REFERENCES

[1] J. NEYMAN, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math France*, Vol. 63 (1935), pp. 246–266.

[2] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc London*, A, Vol. 236 (1937), pp. 333–380

[3] J. NEYMAN, "On a statistical problem arising in routine analyses and in sampling inspections of mass production," *Annals of Math. Stat*, Vol. 12 (1941), pp 46–76.

[4] J. NEYMAN and E S PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil Trans. Roy. Soc London*, A, Vol. 231 (1933), pp 289–337.

[5] J. NEYMAN and E S PEARSON, "Contributions to the theory of testing statistical hypotheses· Part I," *Stat. Res Mem.*, Vol. 1 (1936), pp 1–37.

[6] J. NEYMAN and E. S PEARSON, "Contributions to the theory of testing statistical hypotheses Part II," *Stat. Res Mem.*, Vol. 2 (1938), pp. 25–36.

[7] S BOCHNER, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932

[8] H. CRAMÉR and H. WOLD, "Some theorems on distribution functions," *Jour. London Math. Soc.*, Vol. 11 (1936), pp 290–294.

ON THE PROBLEM OF MULTIPLE MATCHING

By I. L. Battin

Drew University

1. Introduction. The problem of determining the distribution of the number of "hits" or "matchings" under random matching of two decks of cards has received attention from a number of authors within the last few years. In 1934 Chapman [2] considered pairings between two series of t elements each, and later [3] generalized the problem to series of u and $t(\leq u)$ elements respectively. In the same paper he also considered the distribution of the mean number of correct matchings resulting from n independent trials, and gave a method, and tables, for determining the significance of any obtained mean. In 1937 Bartlett [1] considered matchings of two decks of cards, using a number of interesting moment generating functions. In 1937 Huntington [12, 13] gave tables of probabilities for matchings between decks with the compositions (5^5), (4^4), and (3^3), where (s^t) denotes a deck consisting of s of each of t kinds of cards. More generally $(s_1 s_2 \cdots s_t)$ denotes s_1 cards of the first kind, s_2 of the second, etc. Sterne [16] has given the first four moments of the frequency distribution for the (5^5) case and has fitted a Pearson Type I distribution function to the distribution. Sterne obtained his results by considering the probabilities in a 5×5 contingency table. He also considered the 4×4 and 3×3 cases. In 1938 Greville [7] gave a table of the exact probabilities for matchings between two decks of compositions (5^5). Greenwood [4] obtained the variance of the distribution of hits for matchings between two decks having the respective compositions (s^t) and $(s_1 s_2 \cdots s_t)$ with $s_1 + s_2 + \cdots + s_t = st = n$, and where it is not necessary that all the s's should be different from zero. Earlier Wilks [19] had considered the same problem for $t = 5$ and $n = 25$.

In a very interesting paper Olds [15] in 1938 used permanents to express a moment generating function suitable for the problem in question He obtained factorial moments and the first four ordinary moments about the mean, first for two decks with composition (4^2), and then for two decks of composition (s^t). In 1938 Stevens [17] considered a contingency table in connection with matchings between two sets of n objects each, and gave the means, variances, and covariances of the single cell entries and various sub-totals of the cell entries. Stevens [18] also gave a treatment of the problem of matchings between two decks which was based on elementary considerations. In 1940 Greenwood [6] gave the first four moments of the distribution of hits between two decks of any composition whatever, generalizing the problem which had been treated earlier by Olds [15]. Finally in 1941, Greville [8] gave the exact distribution of hits for matchings between two decks of arbitrary composition. He also considered the problem from the standpoint of a contingency table, as had been done earlier by Stevens.

In 1939 Kullback [14] considered matchings between two sequences obtained by drawing at random a single element in turn from each of n urns U_i containing elements of r types E_j in the respective proportions p_{ij}. He showed that if the process of drawing were indefinitely repeated the distribution of hits would be that of a Poisson series.

The work which has been done thus far applies to the problem of matching two decks of cards. In the present paper a method is developed for obtaining the moments of the distribution of hits for matchings between three or more decks of cards of arbitrary composition.

2. Matchings between two Decks of cards In the present paper it will be convenient to take as the point of departure the method used by Wilks [19] in his treatment of the problem of hits occurring under random matching of two decks of 25 cards each, namely a target deck with composition (5^5) and a matching deck with composition (s_i), $i = 1, 2, 3, \cdots, 5$, $\sum_i s_i = 25$ He showed that

$$(1) \quad \phi = \frac{1}{\begin{bmatrix} 25 \\ s_i \end{bmatrix}} (x_1 e^\theta + x_2 + \cdots + x_5)^5 (x_1 + x_2 e^\theta + x_3 + \cdots + x_5)^5$$

$$\cdots (x_1 + x_2 + \cdots + x_5 e^\theta)^5$$

where,

$$\begin{bmatrix} 25 \\ s_i \end{bmatrix} \equiv \frac{25!}{s_1! \, s_2! \cdots s_5!},$$

is a suitable generating function for obtaining the moments of the distribution. In fact, if we define an operator $K_{s_1 s_2 \cdots s_5}$ as

$$(2) \qquad K_{s_1 s_2 \cdots s_5} u \equiv \text{coefficient of } x_1^{s_1} x_2^{s_2} \cdots x_5^{s_5} \text{ in } u,$$

where $u = u(x_1, x_2, \cdots, x_5)$, and if h denotes the number of hits, then for $r = 1, 2, \cdots, 5$,

$$(3) \qquad P(h = r) = \text{coefficient of } e^{r\theta} \text{ in } K_{s_1 s_2 \cdots s_5} \phi$$

And it is readily seen that

$$(4) \qquad E(h^p) = K_{s_1 s_2 \cdots s_5} \left. \frac{\partial^p \phi}{\partial \theta^p} \right|_{\theta = 0}.$$

Wilks' ϕ function involves a particular order for the target deck. If we are to generalize and obtain moments for matchings between more than two decks, it is obvious that we must devise a procedure which will, in the case of two decks, be perfectly symmetrical and not require that one deck be given a preferred status. In the case of two decks this is readily accomplished by the use of Kronecker deltas, and in the case of three or more decks by the use of obvious generalizations of these deltas.

For two decks of 25 cards each with compositions (5^5) we need only let

$$\phi = (x_i\, y_j\, e^{\delta_{ij}\theta})^{25} \equiv \left(\sum_{i,j=1}^{5} x_i\, y_j\, e^{\delta_{ij}\theta}\right)^{25} \tag{5}$$

where $\delta_{ii} = 1$; $\delta_{ij} = 0$, $i \neq j$.
Then, if

(6) $K_{n_{11}n_{12}\cdots n_{15}\cdot n_{21}n_{22}\cdots n_{25}}\, u \equiv$ coefficient of $x_1^{n_{11}}x_2^{n_{12}} \cdots x_5^{n_{15}}y_1^{n_{21}}y_2^{n_{22}} \cdots y_5^{n_{25}}$ in n where $u = u(x_1, x_2, \cdots, x_5, y_1, y_2, \cdots, y_5)$, it readily follows that

$$E(h^p) = \frac{K_{55555\cdot 55555} \left.\dfrac{\partial^p \phi}{\partial \theta^p}\right|_{\theta=0}}{K_{55555\cdot 55555}\, \phi \,|_{\theta=0}}. \tag{7}$$

More generally, for two decks of n cards each, the cards being of k types, and the decks having compositions $(n_{11}, n_{12}, \cdots, n_{1k})$, $(n_{21}, n_{22}, \cdots, n_{2k})$ respectively, we let

$$\phi = u^n \equiv (x_i\, y_j\, e^{\delta_{ij}\theta})^n \equiv \left(\sum_{i,j=1}^{k} x_i\, y_j\, e^{\delta_{ij}\theta}\right)^n. \tag{8}$$

The factors of ϕ are in one-to-one correspondence with the n events of dealing a card from each of the two decks. The values which can be assumed by the subscripts i and j are in one-to-one correspondence with the k types of cards. The symbol x_i corresponds to the first deck, y_j to the second, the subscripts i and j corresponding to the different types of cards in each deck. The expansion of ϕ consists of all products which can be formed by choosing one and only one pair $x_\alpha y_\beta$ from each factor of ϕ as a factor of the product. In forming any term of ϕ, choosing $x_\alpha y_\alpha$ from any factor of ϕ corresponds to dealing a card of type α from both decks, and introduces e^θ into the coefficient of the term. Choosing $x_\alpha y_\beta$ from any factor corresponds to dealing a card of type α from the first deck, β from the second. If $\alpha \neq \beta$, then, since $\delta_{ij} = 0$, $i \neq j$, e^θ is not introduced into the coefficient. Therefore in the coefficient of any term of ϕ, e^θ will be raised to a power, say s, which is equal to the number of factors of ϕ from which pairs $x_\alpha y_\alpha$ have been chosen.

The total number of ways in which the term

$$x_1^{n_{11}}x_2^{n_{12}} \cdots x_k^{n_{1k}}y_1^{n_{21}}y_2^{n_{22}} \cdots y_k^{n_{2k}}$$

can arise is equal to the number of ways in which two decks of types (n_{1i}), (n_{2j}) respectively can be dealt, (where $(n_{1i}) \equiv (n_{11}n_{12} \cdots n_{1k})$ and similarly for (n_{2j})). But this is given by

$$\begin{aligned} K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\, \phi\,|_{\theta=0} &= K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}} \left(\sum_{i=1}^{k} x_i\right)^n \left(\sum_{i=1}^{k} y_i\right)^n \\ &= K_{n_{11}n_{12}\cdots n_{1k}} \left(\sum_{i=1}^{k} x_i\right)^n K_{n_{21}n_{22}\cdots n_{2k}} \left(\sum_{i=1}^{k} y_i\right)^n \\ &= \begin{bmatrix} n \\ n_{1i} \end{bmatrix} \begin{bmatrix} n \\ n_{2j} \end{bmatrix}. \end{aligned} \tag{9}$$

The coefficient of $e^{s\theta}$ in $K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi$ is the total number of ways in which the term $x_1^{n_{11}}x_2^{n_{12}}\cdots x_k^{n_{1k}}y_1^{n_{21}}y_2^{n_{22}}\cdots y_k^{n_{2k}}$ can be formed subject to the restriction that pairs $x_i y_j$ with $i = j$ are chosen from s of the factors of ϕ. But this is precisely the number of ways in which the two decks can be dealt so that there will be s hits. Hence if, as above, h is the number of hits, the probability that $h = s$, assuming all permutations in each deck to be equally likely, is given by

$$P(h = s) = \frac{\text{coefficient of } e^{s\theta} \text{ in } K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}. \tag{10}$$

Since this is true for all values of s it follows that

$$E(h^p) = \frac{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\left.\dfrac{\partial^p\phi}{\partial\theta^p}\right|_{\theta=0}}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}. \tag{11}$$

Since

$$\left.\frac{\partial\phi}{\partial\theta}\right|_{\theta=0} = nu^{n-1}\left[\sum_{i,j=1}^{k}\delta_{ij}x_i y_j e^{\delta_{ij}\theta}\right]\Bigg|_{\theta=0} = n\left[\sum_{i=1}^{k}x_i y_i e^{\theta}\right]u^{n-1}\Bigg|_{\theta=0}$$

$$= n\left[\sum_{i=1}^{k}x_i y_i\right]\left(\sum_{i=1}^{k}x_i\right)^{n-1}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$

we have at once

$$E(h) = \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}K_{n_{11}n_{12}\cdots n_{1i-1}(n_{1i}-1)n_{1i+1}\cdots n_{1k}}\left(\sum_{i=1}^{k}x_i\right)^{n-1}$$

$$\cdot K_{n_{21}n_{22}\cdots n_{2i-1}(n_{2i}-1)n_{2i+1}\cdots n_{2k}}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$

$$= \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}\left[\frac{(n-1)!}{n_{11}!\cdots n_{1i-1}!(n_{1i}-1)!n_{1i+1}!\cdots n_{1k}!}\right] \tag{12}$$

$$\cdot\left[\frac{(n-1)!}{n_{21}!\cdots n_{2i-1}!(n_{2i}-1)!n_{2i+1}!\cdots n_{2k}!}\right]$$

$$= \sum_{i=1}^{k}\frac{n_{1i}n_{2i}}{n}.$$

It is an equally straightforward matter to show that

$$E(h^2) = \sum_i\left[\frac{n_{1i}n_{2i}}{n} + \frac{n_{1i}(n_{1i}-1)n_{2i}(n_{2i}-1)}{n(n-1)}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n(n-1)} \tag{13}$$

and that

$$\sigma_h^2 = \sum_i\left[\frac{n_{1i}n_{2i}}{n} - \frac{n_{1i}^2 n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)}n_{2i}^{(2)}}{n^{(2)}}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n^2(n-1)}. \tag{14}$$

Evidently any of the n_{1i} and n_{2j} may be zero, provided only that $\sum_{i=1}^{k} n_{1i} = \sum_{j=1}^{k} n_{2j} = n$. The case of two decks with unequal numbers of cards m and n, $(m < n)$, is readily handled by substituting for the smaller deck one obtained by adding n-m "blank" cards—that is, cards of any type not already appearing in either deck, as indicated by Greville [8], who however obtained his results by considering a preferred order for one of the decks.

EXAMPLE 1. In the case of the decks treated by Wilks [19], $n = 25$, $k = 5$, $n_{1i} = n_{2j} = 5$. Hence from (12)

$$E(h) = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25}\right\} = 5,$$

and from (14)

$$\sigma_h^2 = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25} - \frac{25\cdot 25}{(25)^2} + \frac{5\cdot 4\cdot 5\cdot 4}{25\cdot 24}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{5\cdot 5\cdot 5\cdot 5}{(25)^2 24}$$

$$= \sum_{i=1}^{5} \frac{16}{24} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{1}{24} = 4\frac{1}{6}.$$

EXAMPLE 2. Suppose we have two decks as shown by the scheme

	Type of card					Total of all types
	1	2	3	4	5	
No. in deck A	5	7	8	0	0	20
No. in deck B	0	3	4	6	2	15

Here deck B has five fewer cards than deck A. Hence we must presume that there are six types of cards in all, and that the decks have the respective distributions (578000) and (034625). We then have at once

$$E(h) = \sum_{i=1}^{6} \frac{n_{1i} n_{2i}}{n} = \frac{1}{20}[0 + 3\cdot 7 + 4\cdot 8 + 0 + 0 + 0]$$

$$= 2.65$$

$$\sigma_h^2 = \sum_{i=1}^{6} \left\{\frac{n_{1i} n_{2i}}{n} - \frac{n_{1i}^2 n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)} n_{2i}^{(2)}}{n^{(2)}}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{6} \frac{n_{1i} n_{2i} n_{1j} n_{2j}}{n^2(n-1)}$$

$$= 2.65 - \frac{1}{400}\{3^2\cdot 7^2 + 4^2\cdot 8^2\} + \frac{1}{20\cdot 19}\{3\cdot 2\cdot 7\cdot 6 + 4\cdot 3\cdot 8\cdot 7\}$$

$$+ \frac{1}{400\cdot 19}\{3\cdot 7\ 4\cdot 8 + 4\cdot 8\cdot 3\cdot 7\}$$

3. Matchings between three decks. Let the three decks be of types $(n_{11}n_{12}\cdots n_{1q})$, $(n_{21}n_{22}\cdots n_{2q})$, $(n_{31}n_{32}\cdots n_{3q})$ respectively, with $\sum_{i=1}^{q} n_{1i} = \sum_{j=1}^{q} n_{2j} = \sum_{k=1}^{q} n_{3k} = n$, and consider the function

$$\phi = \left[\sum_{i,j,k=1}^{q} x_i y_j z_k e^{\delta_{ijk}\theta_{123}+\delta_{ij}\theta_{12}+\delta_{ik}\delta_{13}+\delta_{jk}\theta_{23}}\right]^n \equiv u^n, \tag{15}$$

where

$$\delta_{iii} = 1, \qquad \delta_{ijk} = 0 \qquad i,\ j,\ k \text{ not all equal,} \tag{16}$$

and the other deltas are the usual Kronecker symbols

Each factor of ϕ corresponds to one deal from each of the three decks. The symbols x, y, and z correspond respectively to cards in the first, second, and third decks. The subscripts i, j, k, $= 1, 2, \cdots, q$ correspond to the types of cards—there being q distinct types.

Choosing $x_\alpha y_\alpha z_\alpha$ from a factor of ϕ corresponds to a deal in which a card of type α is dealt from all three decks, and introduces $e^{\theta_{123}+\theta_{12}+\theta_{13}+\theta_{23}}$ into the coefficient of the corresponding term in the expansion of ϕ Similarly, choosing $x_\alpha y_\alpha z_\beta$, $\beta \neq \alpha$, corresponds to a hit between the first and second decks, and introduces $e^{\theta_{12}}$ into the coefficient Similarly choosing $x_\alpha y_\beta z_\alpha$ introduces $e^{\theta_{13}}$; $x_\beta y_\alpha z_\alpha$ introduces $e^{\theta_{23}}$. Choosing $x_\alpha y_\beta z_\gamma$, $\alpha \neq \beta \neq \gamma \neq \alpha$ corresponds to a deal with no hits, and introduces no powers of e into the coefficient, since all the δ's are zero.

Let $K_{n_{1i}, n_{2j}, n_{3k}}$ be defined by

$$K_{n_{1i}, n_{2j}, n_{3k}} u \equiv \text{coefficient of } x_1^{n_{11}} \cdots x_q^{n_{1q}} y_1^{n_{21}} \cdots y_q^{n_{2q}} z_1^{n_{31}} \cdots z_q^{n_{3q}} \text{ in } u. \tag{17}$$

Then the coefficient of $e^{h_{123}\theta_{123}}$ in $K_{n_{1i}, n_{2j}, n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=0}$ is the number of ways in which the cards can be dealt so as to yield precisely h_{123} triples, or hits between all three decks. Similarly the coefficient of $e^{h_{12}\theta_{12}}$ in $K_{n_{1i}, n_{2j}, n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=0}$ is the number of ways in which the cards can be dealt so as to yield precisely h_{12} hits between the first and second decks, with corresponding results for the first and third (h_{13}) and second and third (h_{23}) decks.

By the same reasoning as before then, we have

$$E(h_{123}^r) = \frac{K_{n_{1i}, n_{2j}, n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{123}^r}\right|_{\theta's=0}}{K_{n_{1i}, n_{2j}, n_{3k}}\phi \mid_{\theta's=0}}, \tag{18}$$

$$E(h_{12}^r) = \frac{K_{n_{1i}, n_{2j}, n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{12}^r}\right|_{\theta's=0}}{K_{n_{1i}, n_{2j}, n_{3k}}\phi \mid_{\theta's=0}}, \tag{19}$$

with similar results for h_{13} and h_{23}. And it is a straightforward matter to show that

$$(20)\quad E(h_{123}) = n \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right)$$

$$(21)\quad \begin{aligned} E(h_{123}^2) = n \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right) + n(n-1) \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}^{(2)}}{n^{(2)}} \right) \\ + n(n-1) \sum_{i,j=1,(i\neq j)}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i} n_{\alpha j}}{n^{(2)}} \right). \end{aligned}$$

$$(22)\quad E(h_{12}) = \frac{1}{n^2} \sum_{i,k=1}^{q} n_{1i} n_{2i} n_{3k}$$

$$(23)\quad E(h_{13}) = \frac{1}{n^2} \sum_{j,k=1}^{q} n_{1k} n_{2j} n_{3k}$$

$$(24)\quad E(h_{23}) = \frac{1}{n^2} \sum_{i,j=1}^{q} n_{1i} n_{2j} n_{3j}$$

$$(25)\quad \begin{aligned} E(h_{12}^2) = \frac{1}{n^2} \sum_{i,k} n_{1i} n_{2i} n_{3k} + \frac{1}{n^2(n-1)^2} \Big[\sum_{i,k} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k}^{(2)} \\ + \sum_{i,k\neq r} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k} n_{3r} + \sum_{k,\, i\neq l} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k}^{(2)} \\ + \sum_{i\neq l,\, k\neq r} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k} n_{3r} \Big] \end{aligned}$$

with corresponding results for other moments. It is understood each summation index takes values from 1 to q

As before, if the decks do not all have the same total number of cards it is merely necessary to introduce one or more sets of "blank" cards. Thus we would replace decks with the compositions (57800), (03462), (00335) by hypothetical decks (5780000), (0346250), (0033509) and proceed as before.

EXAMPLE 3. For three decks of 25 cards, consisting of five of each of five kinds we have $n = 25$, $n_{\alpha i} = 5$, $\alpha = 1, 2, 3$, $i = 1, 2, \cdots, 5$. Hence

$$E(h_{123}) = 25 \sum_{i=1}^{5} \prod_{\alpha=1}^{3} \frac{5}{25} = 1$$

$$\begin{aligned} E(h_{123}^2) &= 25 \sum_{i=1}^{5} \left(\frac{5}{25}\right)^3 + 25\cdot 24 \sum_{i=1}^{5} \left(\frac{5\cdot 4}{25\cdot 24}\right)^3 + 25\cdot 24 \sum_{\substack{i,j=1 \\ i\neq j}}^{5} \left(\frac{5^2}{25\cdot 24}\right)^3 \\ &= 1\frac{47}{48} \end{aligned}$$

$$\sigma_{h_{123}}^2 = \frac{47}{48}.$$

$$\begin{aligned} E(h_{12}) &= \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3 \\ &= 5 \end{aligned}$$

$$E(h_{12}^2) = \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3 + \frac{1}{(25)^2(24)^2}\Bigg[\sum_{i,k=1}^{5} 5^3 4^3 + \sum_{\substack{i,k,r=1\\ k\neq r}}^{5} 5^4 4^2 + \sum_{\substack{i,l,k=1\\ i\neq l}}^{5} 5^5 4 + \sum_{\substack{i,l,k,r=1\\ i\neq l\\ k\neq r}}^{5} 5^6\Bigg]$$

$$= 29\tfrac{1}{6},$$

$$\sigma_{h_{12}}^2 = 4\tfrac{1}{6}.$$

with similar results for $E(h_{13})$, $E(h_{23})$, $\sigma_{h_{13}}^2$, and $\sigma_{h_{23}}^2$.

4. Generalization to any number of decks. If the moments of the distribution of hits—doubles, triples, quadruples, .. —in matching any number of decks is desired, these can be obtained by using an obvious generalization of (15). Thus for four decks we would define $\delta_{iiii} = 1$, $\delta_{ijkl} = 0$, i, j, k, l not all equal, and use

$$\phi = \left[\sum_{i,j,k,l=1}^{q} x_i y_j z_k w_l e^{\delta_{ijkl}\theta_{1234}+\delta_{ijl}\theta_{123}+\delta_{ijl}\theta_{124}+\cdots+\delta_{ij}\theta_{12}+\cdots+\delta_{kl}\theta_{34}}\right]^n \tag{26}$$

However, it is evident that as the number of decks is increased the summations involved and the manipulation of the (generalized) K operators rapidly become complicated.

5. Application of our moment-generating technique to two-way contingency tables. The moment-generating technique which we have discussed has wider applications than merely to matching problems. As an example of considerable interest we shall consider the contingency problem. Consider the array

$$\begin{array}{c|c} n_{\alpha\beta\cdot} & n_{\alpha\cdot} \\ \hline n_{\cdot\beta} & n \end{array} \qquad \begin{aligned} \alpha &= 1, 2, \cdots, r \\ \beta &= 1, 2, \cdots, s \\ \sum_{\alpha,\beta} n_{\alpha\beta} &= \sum_{\alpha} n_{\alpha\cdot} = \sum_{\beta} n_{\beta\cdot} = n \end{aligned} \tag{27}$$

and also the function

$$\phi = \prod_{\alpha=1}^{r} (x_\beta e^{\theta_{\alpha\beta}})^{n_{\alpha\cdot}} \equiv \prod_{\alpha=1}^{r} \left(\sum_{\beta=1}^{s} x_\beta e^{\theta_{\alpha\beta}}\right)^{n_{\alpha\cdot}}. \tag{28}$$

If i and j are particular values of α and β respectively, then to the i-th row of the array corresponds the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, consisting of $n_{i\cdot}$ identical factors $x_\beta e^{\theta_{i\beta}}$, one such factor corresponding to each of the $n_{i\cdot}$ elements in the row. To the j-th column of the array corresponds the x_j which appears in each of the factors of ϕ. To the ij-th cell of the array corresponds $e^{\theta_{ij}}$ which appears only in the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, and in each of these only as the coefficient of x_j.

The expansion of ϕ consists of all products which can be formed by taking as factors one and only one element $x_\beta e^{\theta_{\alpha\beta}}$ (not summed) from each factor of ϕ. But taking $x_j e^{\theta_{ij}}$ from one of the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$ of ϕ corresponds exactly to putting an element in the ij-th cell of a lattice such as (27). Hence every term in the expansion of ϕ corresponds to a particular distribution in such a lattice. Moreover, all terms of ϕ correspond to distributions in which the row totals are $n_{\alpha\cdot}$, for we must take $n_{i\cdot}$ elements from the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$. Further, those terms in which the x_β appear in the particular product $x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}}$ correspond to distributions in which the column totals are $n_{\cdot 1}, n_{\cdot 2}, \cdots, n_{\cdot s}$, since choosing $n_{\cdot j}$ elements $x_j e^{\theta_{\alpha j}}$ corresponds to putting $n_{\cdot j}$ elements in the j-th column and some row of the lattice.

Expanding ϕ we obtain

$$\phi = \cdots + \left[\sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} \right] x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots \tag{29}$$

where the summation is over all partitions $(n_{\alpha 1} n_{\alpha 2} \cdots n_{\alpha s})$ of the $n_{\alpha\cdot}$ such that $(n_{1\beta} n_{2\beta} \cdots n_{r\beta})$ is also a partition of $n_{\cdot\beta}$. It is clear that since every set of values of the $n_{\alpha\beta}$ subject to the partition restrictions $\sum_\beta n_{\alpha\beta} = n_{\alpha\cdot}$, $\sum_\alpha n_{\alpha\beta} = n_{\cdot\beta}$ corresponds to a particular distribution of n elements in the lattice (27), every particular product $\prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$ corresponds to such a distribution, and represents the number of ways in which it can arise. Further, the total coefficient displayed (29), namely $\sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$, represents the total number of ways in which distributions with row totals $n_{\alpha\cdot}$ and column totals $n_{\cdot\beta}$ can arise. Setting all the $\theta_{\alpha\beta} = 0$ we readily find

$$\begin{aligned} \sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} &= K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} \phi \,|_{\theta_{\alpha\beta}=0} = K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} (x_1 + x_2 + \cdots + x_s)^n \\ &= \begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}. \end{aligned} \tag{30}$$

Hence the probability of any particular distribution $\| n_{\alpha\beta} \|$ with fixed row totals $n_{\alpha\cdot}$ and fixed column totals $n_{\cdot\beta}$ is

$$P(\| n_{\alpha\beta} \| \,|\, n_{\alpha\cdot}, n_{\cdot\beta}) = \frac{\prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}}{\begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}}. \tag{31}$$

Moments of the n_{ij}. Consider now the result of differentiating ϕ with respect to a particular $\theta_{\alpha\beta}$, say θ_{ij}. We obtain

$$\frac{\partial\phi}{\partial\theta_{ij}} = \cdots + \sum\nolimits_a n_{ij} \prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots \tag{32}$$

where $\sum_a$ denotes summation over indices such that $\sum_\alpha n_{\alpha\beta} = n_{.\beta}$, $\sum_{\alpha \neq i} n_{\alpha j} + n_{ij} = n_{.j}$ $(\beta \neq j)$. Now $n_{ij} \leq \min (n_{i.}, n_{.j})$, but also n_{ij} can never be less than $n_{.j} - (n - n_{i.})$. For $n_{.j} = n_{ij} + \sum_{\alpha \neq i} n_{\alpha j}$. Since the maximum value of $n_{\alpha j} \leq n_{\alpha.}$, the maximum value of $\sum_{\alpha \neq i} n_{\alpha j} \leq \sum_{\alpha \neq i} n_{\alpha.}$. Hence

$$n_{ij} = n_{.j} - \sum_{\alpha \neq i} n_{\alpha j} \geq n_{.j} - \sum_{\alpha \neq i} n_{\alpha.} = n_{.j} - (n - n_{i.}).$$

Therefore

$$\max (0, n_{.j} - n + n_{i.}) \leq n_{ij} \leq \min (n_{i.}, n_{.j}).$$

Accordingly, combining all the terms of (32) in which n_{ij} has a particular value, γ, we have

$$\frac{\partial \phi}{\partial \theta_{ij}} = \cdots + \sum_{\gamma = \max (0, n_{.j} - n + n_{i.})}^{\min (n_{i.}, n_{.j})} \gamma \sum^*_{\sum_\alpha n_{\alpha\beta} = n_{.\beta}} \prod_\alpha^* \begin{bmatrix} n_{\alpha.} \\ n_{\alpha\beta} \end{bmatrix} \tag{33}$$

$$\cdot e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{.1}} x_2^{n_{.2}} \cdots x_s^{n_{.s}} + \cdots.$$

where Σ^* denotes summation and Π^* multiplication with $n_{ij} = \gamma$.

Since $\sum^*_{\sum_\alpha n_{\alpha\beta} = n_{.\beta}} \prod_\alpha^* \begin{bmatrix} n_{\alpha.} \\ n_{\alpha\beta} \end{bmatrix}$ is precisely the number of distributions $\| n_{\alpha\beta} \|$ for which $n_{ij} = \gamma$, it follows that

$$E(n_{ij} \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial \phi}{\partial \theta_{ij}} \right|_{\theta_{\alpha\beta} = 0}. \tag{34}$$

Similarly it follows that

$$E(n_{ij}^p \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^p \phi}{\partial \theta_{ij}^p} \right|_{\theta_{\alpha\beta} = 0} \tag{35}$$

$$E(n_{ij}^p n_{kl}^q \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^{p+q} \phi}{\partial \theta_{ij}^p \, \partial \theta_{kl}^q} \right|_{\theta_{\alpha\beta} = 0} \tag{36}$$

where we may have $i = k$ or $i \neq k$, and $j = l$ or $j \neq l$.

By straightforward differentiation and reduction we find that for the array (27) with given marginal totals $n_{\alpha.}$, $n_{.\beta}$

$$E(n_{ij}) = \frac{n_{i.} n_{.j}}{n} \tag{37}$$

$$E(n_{ij}^2) = \frac{n_{i.}^{(2)} n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.} n_{.j}}{n} \tag{38}$$

$$(39)\qquad \sigma^2_{n_{ij}} = \frac{[n^2 - n(n_{i.} + n_{.j}) + n_{i.}n_{.j}]n_{i.}n_{.j}}{n^2(n-1)}$$

$$(40)\qquad E(n^3_{ij}) = \frac{n^{(3)}_{i.}\,n^{(3)}_{.j}}{n^{(3)}} + 3\,\frac{n^{(2)}_{i.}\,n^{(2)}_{.j}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n}$$

$$(41)\qquad E(n^4_{ij}) = \frac{n^{(4)}_{i.}\,n^{(4)}_{.j}}{n^{(4)}} + 6\,\frac{n^{(3)}_{i.}\,n^{(3)}_{.j}}{n^{(3)}} + 7\,\frac{n^{(2)}_{i.}\,n^{(2)}_{.j}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n},$$

and if i and k, j and l are distinct

$$(42)\quad E(n^2_{ij}n^2_{kj}) = \frac{n^{(2)}_{i.}\,n^{(2)}_{k.}\,n^{(4)}_{.j}}{n^{(4)}} + (n^{(2)}_{i.}\,n_{k.} + n_{i.}n^{(2)}_{k.})\,\frac{n^{(3)}_{.j}}{n^{(3)}} + \frac{n_{i.}n_{k.}n^{(2)}_{.j}}{n^{(2)}}$$

$$(43)\quad E(n^2_{ij}n^2_{il}) = \frac{n^{(4)}_{i.}\,n^{(2)}_{.j}\,n^{(2)}_{.l}}{n^{(4)}} + (n^{(2)}_{.j}\,n_{.l} + n_{.j}n^{(2)}_{.l})\,\frac{n^{(3)}_{i.}}{n^{(3)}} + \frac{n^{(2)}_{i.}\,n_{.j}n_{.l}}{n^{(2)}}$$

$$(44)\quad \begin{aligned} E(n^2_{ij}\,n^2_{kl}) = {} & \frac{n^{(2)}_{i.}\,n^{(2)}_{k.}\,n^{(2)}_{.j}\,n^{(2)}_{.l}}{n^{(4)}} + \frac{n^{(2)}_{i.}\,n_{k.}n^{(2)}_{.j}\,n_{.l}}{n^{(3)}} \\ & + \frac{n_{i.}n^{(2)}_{k.}\,n_{.j}n^{(2)}_{.l}}{n^{(3)}} + \frac{n_{i.}n_{k.}n_{.j}n_{.l}}{n^{(2)}} \end{aligned}$$

Moments of the distribution of Chi Square. For the array (27)

$$(45)\qquad \begin{aligned} \chi^2 &= \sum_{\alpha,\beta} \frac{\left(n_{\alpha\beta} - \dfrac{n_{\alpha.}n_{.\beta}}{n}\right)^2}{\dfrac{n_{\alpha.}n_{.\beta}}{n}} \\ &= \sum_{\alpha,\beta}\left[\frac{n}{n_{\alpha.}n_{.\beta}}\,n^2_{\alpha\beta} - 2n_{\alpha\beta} + \frac{n_{\alpha.}n_{.\beta}}{n}\right]. \end{aligned}$$

Hence, using the above results we can, theoretically, find all the moments of the exact distribution of χ^2. It is not difficult to show that

$$(46)\qquad E(\chi^2) = \frac{n}{n-1}(r-1)(s-1).$$

The value of $E[(\chi^2)^2]$ and the variance of χ^2 were found by straightforward application of our methods and the results agreed with those given by Haldane [10].

The writer is indebted to Professor Wilks for helpful criticisms and suggestions.

REFERENCES

[1] M S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc.*, A, Vol. 160 (1937), pp. 268–282.

[2] Dwight Chapman, "The statistics of the method of correct matchings," *Amer. Jour. Psych.*, Vol. 46 (1934), pp. 287–298

[3] Dwight Chapman, "The generalized problem of correct matchings," *Annals of Math. Stat.*, Vol. 6 (1935), pp 85–95.

[4] J. A GREENWOOD, "Variance of a general matching problem," *Annals of Math. Stat.*, Vol. 9 (1938), pp 56–59.

[5] J. A GREENWOOD, "Variance of the ESP call series," *Jour. of Parapsychology*, Vol 2 (1938), pp. 60–64

[6] J A. GREENWOOD, "The first four moments of a general matching problem," *Annals of Eug.*, Vol 10 (1940), pp. 290–292.

[7] T. N. E. GREVILLE, "Exact probabilities for the matching hypothesis," *Jour. of Parapsychology*, Vol. 2 (1938), pp. 55–59

[8] T. N. E. GREVILLE, "The frequency distribution of a general matching problem," *Annals of Math. Stat.*, Vol. 12 (1941), pp 350–354

[9] J. B S HALDANE, "The mean and variance of Chi square, when used as a test of homogeneity, when expectations are small," *Biometrika*, Vol 31 (1940)

[10] J. B S. HALDANE, "The first six moments of Chi-square for an n-fold table with n degrees of freedom when some expectations are small," *Biometrika*, Vol 29 (1939), p. 389

[11] J. B. S HALDANE, "The exact value of the moments of the distribution of chi-square, used as a test of goodness of fit, when the expectations are small," *Biometrika*, Vol. 29 (1939), p. 133.

[12] E. V. HUNTINGTON, "A rating table for card matching experiments," *Jour. of Parapsychology*, Vol. 4 (1937), pp 292–294

[13] E V HUNTINGTON, "Exact probabilities in certain card-matching problems," *Science*, Vol. 86 (1937), pp. 499–500

[14] SOLOMON KULLBACK, "Note on a matching problem," *Annals of Math Stat*, Vol 10 (1939), pp. 77–80.

[15] E G. OLDS, "A moment-generating function which is useful in solving certain matching problems," *Bull Amer. Math. Soc.*, Vol. 44 (1938), pp. 407–413.

[16] T. E. STERNE, "The solution of a problem in probability," *Science*, Vol 86, (1937) pp 500–501.

[17] W L STEVENS, "Distribution of entries in a contingency table," *Annals of Eugenics*, Vol. 8 (1938), pp. 238–244.

[18] W. L. STEVENS, "Tests of significance for extra sensory perception data," Psychological Review, Vol. 46 (1938), pp 142–150.

[19] S. S. WILKS, "Statistical aspects of experiments in telepathy," a lecture delivered to the Galois Institute of Mathematics, Long Island University, December 4, 1937

ON THE CHOICE OF THE NUMBER OF CLASS INTERVALS IN THE APPLICATION OF THE CHI SQUARE TEST

By H. B. Mann and A. Wald[1]

Columbia University

Introduction. To test whether a sample has been drawn from a population with a specified probability distribution, the range of the variable is divided into a number of class intervals and the statistic,

$$\sum_{i=1}^{i=k} \frac{(\alpha_i - Np_i)^2}{Np_i} = \chi^2, \tag{1}$$

computed. In (1) k is the number of class intervals, α_i the number of observations in the ith class, p_i the probability that an observation falls into the ith class (calculated under the hypothesis to be tested). It is known that under the null hypothesis (hypothesis to be tested) the statistic (1) has asymptotically the chi-square distribution with $k - 1$ degrees of freedom, when each Np_i is large. To test the null hypothesis the upper tail of the chi-square distribution is used as a critical region.

In the literature only rules of thumb are found as to the choice of the number and lengths of the class intervals. It is the purpose of this paper to formulate principles for this choice and to determine the number and lengths of the class intervals according to these principles.

If a choice is made as to the number of class intervals it is always possible to find alternative hypotheses with class probabilities equal to the class probabilities under the null hypothesis. The least upper bound of the "distances" of such alternative distributions from the null hypothesis distribution can evidently be minimized by making the class probabilities under the null hypothesis equal to each other. By the distance of two distribution functions we mean the least upper bound of the absolute value of the difference of the two cumulative distribution functions We have therefore based this paper on a procedure by which the lengths of the class intervals are determined so that the probability of each class under the null hypothesis is equal to $1/k$ where k is the number of class intervals.[2]

Let $C(\Delta)$ be the class of alternative distributions with a distance $\geqslant \Delta$ from the null hypothesis. Let $f(N, k, \Delta)$ be the greatest lower bound of the power of the chi-square test with sample size N and number of class intervals k with respect to alternatives in $C(\Delta)$. The maximum of $f(N, k, \Delta)$ with respect to k is a function $\Phi(N, \Delta)$ of N and Δ. It is most desirable to maximize $f(N, k, \Delta)$ for

[1] Research under a grant in aid from the Carnegie Corporation of New York

[2] This procedure was first used by H Hotelling. "The consistency and ultimate distribution of optimum statistics," *Trans. Am. Math. Soc*, Vol. 32, pp 851.) It has been advocated by E. J. Gumbel in a paper which will appear shortly.

We now assume that N is so large that the joint distribution of the z_i is sufficiently well approximated by a multivariate normal distribution. Then

$E(z_i^2 z_j) = 0, \quad E(z_i^4) = 3[E(z_i^2)]^2, \quad E(z_i^2 z_j^2) = E(z_i^2)E(z_j^2) + 2[E(z_i z_j)]^2$ for $i \neq j$.

We have the well known relations

$$E(z_i^2) = E(\alpha_i^2) - N^2 p_i^2 = Np_i(1 - p_i),$$
$$E(z_i z_j) = E(\alpha_i \alpha_j) - N^2 p_i p_j = -Np_i p_j .$$

Using the above equations we obtain

$$\sigma^2_{\chi'^2} = \frac{k^2}{N^2}\left\{E\left(\sum_{i=1}^{i=k} z_i^2\right)^2 - \left(E\sum_{i=1}^{i=k} z_i^2\right)^2 + 4E\left(\sum_{i=1}^{i=k} z_i \nu_i\right)^2\right\},$$

$$E\left(\sum_{i=1}^{i=k} z_i^2\right)^2 - \left(E\sum_{i=1}^{i=k} z_i^2\right)^2$$
$$= N^2\left\{3\sum_{i=1}^{i=k} p_i^2(1-p_i)^2 + \sum_{i \neq j}[p_i p_j(1-p_i)(1-p_j) + 2p_i^2 p_j^2] - \left[\sum_{i=1}^{i=k} p_i(1-p_i)\right]^2\right\}$$
$$= 2N^2\left[\sum_{i=1}^{i=k} p_i^2(1-p_i)^2 + \sum_{i \neq j} p_i^2 p_j^2\right]$$
$$= 2N^2\left[\sum_{i=1}^{i=k} p_i^2 - 2\sum_{i=1}^{i=k} p_i^3 + \left(\sum_{i=1}^{i=k} p_i^2\right)^2\right].$$

Further

$$E\left(\sum_{i=1}^{i=k} z_i \nu_i\right)^2 = E\left(\sum_{i=1}^{i=k} z_i^2 \nu_i^2\right) + E\left(\sum_{i \neq j} z_i z_j \nu_i \nu_j\right)$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i(1-p_i)\left(p_i - \frac{1}{k}\right)^2 - \sum_{i \neq j} p_i p_j\left(p_i - \frac{1}{k}\right)\left(p_j - \frac{1}{k}\right)\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i\left(p_i - \frac{1}{k}\right)^2 - \left[\sum_{i=1}^{i=k} p_i\left(p_i - \frac{1}{k}\right)\right]^2\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i^3 - \frac{2}{k}\sum_{i=1}^{i=k} p_i^2 + \frac{1}{k^2} - \left[\sum_{i=1}^{i=k} p_i^2 - \frac{1}{k}\right]^2\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i^3 - \left(\sum p_i^2\right)^2\right].$$

Substituting this into the formula for $\sigma^2_{\chi'^2}$ we finally obtain

$$\sigma^2_{\chi'^2} = 2k^2\left\{\sum_{i=1}^{i=k} p_i^2 + 2(N-1)\sum_{i=1}^{i=k} p_i^3 - (2N-1)\left(\sum_{i=1}^{i=k} p_i^2\right)^2\right\}. \tag{4}$$

2. The Taylor expansion of the power. Let C be determined so that the probability under the null hypothesis that $\sum_{i=1}^{i=k} x_i^2 \geqslant C$ is equal to the size λ_0 of

the critical region. Let $P\left(\sum_{i=1}^{i=k} x_i^2 \geqslant C\right)$ be the probability under the alternative hypothesis that $\sum_{i=1}^{i=k} x_i^2 \geqslant C$. Then the power P is given by

$$P\left(\sum_{i=1}^{i=k} x_i^2 \geqslant C\right), \tag{5}$$

where

$$x_i = \frac{\alpha_i - \frac{N}{k}}{\sqrt{\frac{N}{k}}}.$$

Hence

$$\sum_{i=1}^{i=k} x_i^2 = \frac{k}{N}\left(\sum_{i=1}^{i=k} \alpha_i^2 - \frac{N^2}{k}\right),$$

and (5) can be written in the form

$$P\left(\sum_{i=1}^{i=k} \alpha_i^2 \geqslant C'\right) \tag{6}$$

where C' is a certain function of N and k. Let $\delta_i = p_i - \frac{1}{k}$, where p_i is the probability of the ith class interval under the alternative hypothesis.

Expanding P into a power series we obtain (in this and the following derivations, we take all partial differential quotients at the point $\delta_1 = \delta_2 = \cdots = \delta_k = 0$)

$$P = \lambda_0 + \sum_{i=1}^{i=k} \delta_i \frac{\partial P}{\partial \delta_i} + \frac{1}{2}\left\{\sum_{i=1}^{i=k} \delta_i^2 \frac{\partial^2 P}{\partial \delta_i^2} + \sum_{i \neq j} \delta_i \delta_j \frac{\partial^2 P}{\partial \delta_i \partial \delta_j}\right\} + \cdots.$$

Since P is a symmetric function of the δ_i we have for $\delta_1 = \delta_2 = \cdots = \delta_k = 0$

$$\frac{\partial^2 P}{\partial \delta_i^2} = \frac{\partial^2 P}{\partial \delta_1^2}, \qquad \frac{\partial^2 P}{\partial \delta_i \partial \delta_j} = \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2} \qquad \text{for } i \neq j.$$

Furthermore $\sum_{i=1}^{i=k} \delta_i = 0$. Therefore

$$P = \lambda_0 + \frac{1}{2}\left\{\frac{\partial^2 P}{\partial \delta_1^2} \sum_{i=1}^{i=k} \delta_i^2 + \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2} \sum_{i \neq j} \delta_i \delta_j\right\} + \cdots.$$

We shall first show that the terms of second order are always positive. This shows that the test is unbiased and justifies again the choice of equal class probabilities under the null hypothesis since this assures unbiasedness and mini-

mizes among all unbiased tests the g.l.b. of the distances of such alternatives whose power is equal to the size of the critical region.
The power is given by

$$P = \sum_{\alpha_1^2+\alpha_2^2+\cdots\alpha_k^2 \geq C'} \frac{N!}{\alpha_1!\alpha_2! \cdots \alpha_k!} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}.$$

Since $\sum_{i=1}^{i=k} \delta_i^2 = -\sum_{i \neq j} \delta_i \delta_j$ we obtain for the second order terms

$$\begin{aligned} \frac{\partial^2 P}{\partial \delta_1^2} \sum_{i=1}^{i=k} \delta_i^2 + \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2} \sum_{i \neq j} \delta_i \delta_j &= \left(\frac{\partial^2 P}{\partial \delta_1^2} - \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2}\right) \sum_{i=1}^{i=k} \delta_i^2 \\ &= \sum_{\alpha_1^2+\alpha_2^2+\cdots+\alpha_k^2 \geq C'} (\alpha_1^2 - \alpha_1 - \alpha_1 \alpha_2) p(\alpha_1, \alpha_2 \cdots \alpha_k) \sum_{i=1}^{i=k} \delta_i^2 \end{aligned} \tag{7}$$

where

$$p(\alpha_1, \cdots \alpha_k) = \frac{N!}{\alpha_1!\alpha_2! \cdots \alpha_k!} \frac{1}{k^N}.$$

In the following derivation extend all sums if not otherwise stated over all terms for which $\sum_{i=1}^{i=k} \alpha_i^2 \geqslant C'$ and use the relation $\sum_{i=1}^{i=k} \alpha_i = N$. We have because of the symmetry

$$\sum \alpha_1 p(\alpha_1, \alpha_2, \cdots \alpha_k) = \frac{N}{k} \sum p(\alpha_1, \alpha_2 \cdots \alpha_k) = \frac{N}{k} \lambda_0,$$

$$\begin{aligned} \sum \alpha_1 \alpha_2 p(\alpha_1, \alpha_2, \cdots \alpha_k) &= \frac{1}{k(k-1)} \sum \left(N^2 - \sum_{i=1}^{i=k} \alpha_i^2\right) p(\alpha_1, \alpha_2, \cdots \alpha_k) \\ &= \frac{N^2 \lambda_0}{k(k-1)} - \frac{1}{k-1} \sum \alpha_1^2 p(\alpha_1, \alpha_2, \cdots \alpha_k). \end{aligned}$$

Hence the coefficient of the second order term becomes

$$\begin{aligned} &\frac{k}{k-1} \sum \alpha_1^2 p(\alpha_1, \alpha_2, \cdots \alpha_k) - \frac{N}{k} \lambda_0 - \frac{N^2}{k(k-1)} \lambda_0 \\ &\qquad = \frac{1}{k-1} \sum \sum_{i=1}^{i=k} \alpha_i^2 p(\alpha_1, \alpha_2, \cdots \alpha_k) - \frac{N}{k} \lambda_0 - \frac{N^2}{k(k-1)} \lambda_0. \end{aligned}$$

But

$$\frac{\sum \sum_{i=1}^{i=k} \alpha_i^2 p(\alpha_1, \alpha_2, \cdots \alpha_k)}{\lambda_0} > E\left(\sum_{i=1}^{i=k} \alpha_i^2\right),$$

since the conditional mean for values of $\sum_{i=1}^{i=k} \alpha_i^2 \geqslant C'$ must be larger than the

mean of all values of $\sum_{i=1}^{i=k} \alpha_i^2$. Since $E\left(\sum_{i=1}^{i=k} \alpha_i^2\right) = \frac{N^2}{k} - \frac{N}{k} + N$, we obtain

$$\frac{1}{k-1} \sum \sum_{i=1}^{i=k} \alpha_i^2 p(\alpha_1, \alpha_2 \cdots \alpha_k)$$

$$> \frac{\lambda_0}{k-1}\left(\frac{N^2}{k} + \frac{N(k-1)}{k}\right) = \lambda_0\left(\frac{N^2}{k(k-1)} + \frac{N}{k}\right)$$

and hence the coefficient of $\sum_{i=1}^{i=k} \delta_i^2$ is larger than 0.

To prove Theorem 1, we will have to determine the alternative distribution for which $\sum_{i=1}^{i=k} \delta_i^2$ becomes a minimum subject to the condition that the distance from the null hypothesis should be greater than or equal to a given Δ.

Hence we have to find a distribution function $F(x)$ such that $| F(x) - x | \geqslant \Delta$ for at least one value x and $\sum_{i=1}^{i=k} \delta_i^2 = \sum_{i=1}^{i=k} \left(p_i - \frac{1}{k}\right)^2 = \sum_{i=1}^{i=k} p_i^2 - \frac{1}{k}$ is a minimum where $p_i = F\left(\frac{i}{k}\right) - F\left(\frac{i-1}{k}\right)$. Instead of minimizing $\sum_{i=1}^{i=k} \delta_i^2$ we may minimize $\sum_{i=1}^{i=k} p_i^2$, since the two expressions differ merely by a constant. There will be two different solutions for $F(x)$ depending on whether $F(x) - x \geqslant \Delta$ or $F(x) - x \leqslant -\Delta$ for at least one value x. Because of symmetry we restrict ourselves to the case in which $F(x) - x \geqslant \Delta$ for at least one value of x.

Let a be a value for which $F(a) - a \geqslant \Delta$ and suppose that

$$\frac{l-1}{k} < a \leqslant \frac{l}{k}$$

then

$$F(a) \geqslant a + \Delta,$$

$$F\left(\frac{l}{k}\right) = \frac{l}{k} + \epsilon.$$

We prove first

$$\epsilon \geqslant \Delta - \frac{1}{k}.$$

PROOF: Since $F\left(\frac{l}{k}\right) - F(a) \geqslant 0$ we have

$$F\left(\frac{l}{k}\right) = F(a) + F\left(\frac{l}{k}\right) - F(a) \geqslant a + \Delta$$

such values of Δ for which $\Phi(N, \Delta)$ is neither too large nor too small and in this paper we propose to determine Δ so that $\Phi(N, \Delta)$ is equal to $\frac{1}{2}$.

Hence we introduce the following definitions:

DEFINITION 1. *A positive integer k is called best with respect to the number of observations N if there exists a Δ such that $f(N, k, \Delta) = \frac{1}{2}$ and $f(N, k', \Delta) \leqslant \frac{1}{2}$ for any positive integer k'.*

DEFINITION 2 *A positive integer k is called ϵ-best $(0 \leqslant \epsilon \leqslant 1)$ with respect to the number of observations N if ϵ is the smallest number in the interval $[0, 1]$ for which the following condition is fulfilled. There exists a Δ such that $f(N, k, \Delta) \geqslant \frac{1}{2} - \epsilon$ and $f(N, k', \Delta) \leqslant \frac{1}{2} + \epsilon$ for any positive integer k'.*

It is obvious that an ϵ-best k is a best k if $\epsilon = 0$. If ϵ is very small an ϵ-best k is for all practical purposes equivalent to a best k.

Since $f(N, k, \Delta)$ is a continuous function of Δ it is easy to see that for any pair of positive integers k and N there exists exactly one value ϵ such that k is ϵ-best with respect to the number of observations N. Since the value of this ϵ is a function of k and N we will denote it by $\epsilon(k, N)$.

DEFINITION 3. *A sequence $\{k_N\}$ of positive integers is called best in the limit if $\lim_{N=\infty} \epsilon(k_N, N) = 0$.*

In this paper the following theorem is proved:

THEOREM 1. *Let $k_N = 4\sqrt[5]{\dfrac{2(N-1)^2}{c^2}}$ where c is determined so that $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_c^\infty e^{-x^2/2}\,dx$ is equal to the size of the critical region (probability of the critical region under the null hypothesis) then the sequence $\{k_N\}$ is best in the limit. Furthermore $\lim_{N=\infty} f(N, k_N, \Delta_N) = \frac{1}{2}$ for $\Delta_N = \dfrac{5}{k_N} - \dfrac{4}{k_N^2}$.*

It is further shown that for $N \geqslant 450$, if the 5% level of significance is used, and for $N \geqslant 300$, if the 1% level of significance is used, the value of $\epsilon(k_N, N)$ is small so that for practical purposes k_N can be considered as a best k. The authors are convinced although no rigorous proof has been given that $\epsilon(k_N, N)$ is quite small for $N \geqslant 200$ and is very likely to be small even for considerably lower values of N.

1. Mean value and standard deviation of the statistic under alternative hypotheses. It is well known that every continuous distribution can by a simple transformation be transformed into a rectangular distribution with range [0, 1]. We may therefore for convenience assume that the hypothesis to be tested is that of a rectangular distribution with the range [0, 1]. Moreover as mentioned earlier we assume that a procedure is chosen by which the class probabilities under the null hypothesis are equal to each other.

The statistic whose mean value and standard deviation is to be determined is

$$\sum_{i=1}^{i=k} x_i^2 = \chi'^2 \quad \text{where} \quad x_i = \sqrt{\frac{k}{N}}\left(\alpha_i - \frac{N}{k}\right).$$

Let p_i be the probability under the alternative hypothesis that one observation will fall into the ith class. The probability of obtaining certain specified values $\alpha_1, \alpha_2, \cdots, \alpha_k$ is given by

$$f(\alpha_1, \alpha_2, \cdots \alpha_k) = \frac{N!}{\alpha_1!\alpha_2! \cdots \alpha_k!} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}.$$

Since $\sum_{i=1}^{i=k} \alpha_i = N$ we have

$$\sum_{i=1}^{i=k} x_i^2 = \frac{k}{N} \sum_{i=1}^{i=k} \alpha_i^2 - N.$$

We consider the function

$$(p_1e^{t_1} + p_2e^{t_2} + \cdots p_ke^{t_k})^N = \Sigma f(\alpha_1, \alpha_2, \cdots \alpha_k)e^{\alpha_1 t_1 + \alpha_2 t_2 + \cdots \alpha_k t_k}.$$

Differentiating twice and then setting $t_i = 0$ for $i = 1, 2, \cdots k$ we obtain

(2) $N(N-1)p_i^2 + Np_i = E(\alpha_i^2), \quad N(N-1)p_ip_j = E(\alpha_i\alpha_j)$ for $i \neq j$.

Hence

$$E\left(\sum_{i=1}^{i=k} a_i^2\right) = N(N-1) \sum_{i=1}^{i=k} p_i^2 + N,$$

and

$$E(\chi'^2) = k(N-1) \sum_{i=1}^{i=k} p_i^2 + k - N. \tag{3}$$

To compute the standard deviation of χ'^2 we put

$$\mu_i = \left(Np_i - \frac{N}{k}\right)\sqrt{\frac{k}{N}} = \sqrt{Nk}\left(p_i - \frac{1}{k}\right),$$

$$y_i = (\alpha_i - Np_i)\sqrt{\frac{k}{N}} \qquad \text{hence} \qquad y_i = x_i - \mu_i, \qquad E(y_i) = 0.$$

We have

$$\sigma^2_{\chi'^2} = E\left[\sum_{i=1}^{i=k} (y_i + \mu_i)^2 - E\left(\sum_{i=1}^{i=k} (y_i + \mu_i)^2\right)\right]^2$$
$$= E\left(\sum_{i=1}^{i=k} y_i^2 + 2\sum_{i=1}^{i=k} y_i\mu_i - E\left(\sum_{i=1}^{i=k} y_i^2\right)\right)^2.$$

Let

$$\sqrt{\frac{N}{k}}\, y_i = z_i, \qquad \sqrt{\frac{N}{k}}\, \mu_i = \nu_i;$$

then

$$\nu_i = N\left(p_i - \frac{1}{k}\right), \qquad z_i = \alpha_i - Np_i.$$

and

$$\epsilon = F\left(\frac{l}{k}\right) - \frac{l}{k} \geqslant a + \Delta - \frac{l}{k} \geqslant \frac{l-1}{k} + \Delta - \frac{l}{k} = \Delta - \frac{1}{k}.$$

If $\Delta \leqslant \frac{1}{k}$ we can always find a distribution function in $C(\Delta)$ for which $p_i = \frac{1}{k}$. Hence we consider only the case $k > \frac{1}{\Delta}$. We must minimize $\sum_{i=1}^{i=k} p_i^2$ under the condition $\sum_{i=1}^{i=l} p_i = \frac{l}{k} + \epsilon, \sum_{i=l+1}^{i=k} p_i = \frac{k-l}{k} - \epsilon$. We therefore minimize

$$\Phi = \sum_{i=1}^{i=k} p_i^2 - 2\lambda_1 \sum_{i=1}^{i=l} p_i - 2\lambda_2 \sum_{i=l+1}^{i=k} p_i .$$

This leads to

$$p_i = \begin{cases} \dfrac{1}{k} + \dfrac{\epsilon}{l} & \text{for} \quad i = 1, \cdots l \\[2ex] \dfrac{1}{k} - \dfrac{\epsilon}{k-l} & \text{for} \quad i = (l+1), \cdots k. \end{cases}$$

We then have

$$\sum_{i=1}^{i=k} p_i^2 = l\left(\frac{1}{k} + \frac{\epsilon}{l}\right)^2 + (k-l)\left(\frac{1}{k} - \frac{\epsilon}{k-l}\right)^2 = \frac{1}{k} + \frac{\epsilon^2 k}{l(k-l)}.$$

This is smallest if $\epsilon = \Delta - \frac{1}{k}$ and $l = \frac{k}{2}$. The following discontinuous distribution function gives these values for ϵ, l and p_i and has the distance Δ from the rectangular distribution.

$$\begin{aligned} &F(x) = x\left[1 + 2\left(\Delta - \frac{1}{k}\right)\right] && \text{for} \quad 0 \leqslant x \leqslant \frac{1}{2} - \frac{1}{k}, \\ &F(x) = \frac{1}{2} + \Delta - \frac{1}{k} && \text{for} \quad \frac{1}{2} - \frac{1}{k} < x \leqslant \frac{1}{2}, \\ (8) \quad &F(x) = x\left[1 - 2\left(\Delta - \frac{1}{k}\right)\right] + 2\left(\Delta - \frac{1}{k}\right) && \text{for} \quad \frac{1}{2} \leqslant x \leqslant 1, \\ &F(x) = 0 && \text{for} \quad 0 \leqslant x, \\ &F(x) = 1 && \text{for} \quad x \geqslant 1. \end{aligned}$$

3. Solution for large N. Denote by $F(\Delta, k)$ the distribution function (8) of $C(\Delta)$ which makes $\sum_{i=1}^{i=k} \delta_i^2$ a minimum if the test is made with k class intervals.

Assume that k is large enough that χ'^2 can be taken as normally distributed. The power of the test is then given by

$$\frac{1}{\sqrt{2\pi}\,\sigma'}\int_{(k-1)+c\sqrt{2(k-1)}}^{\infty} e^{-\frac{1}{2\sigma'^2}\left(\sum_{i=1}^{i=k} x_i^2 - E\left(\sum_{i=1}^{i=k} x_i^2\right)\right)^2} d\left(\sum_{i=1}^{i=k} x_i^2\right) \tag{9}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{k-1-E\left(\sum_{i=1}^{i=k} x_i^2\right)+c\sqrt{2(k-1)}}{\sigma'}}^{\infty} e^{-\frac{1}{2}y^2}\,dy,$$

where σ' is the standard deviation of $\sum_{i=1}^{i=k} x_i^2$ and c is determined so that $\frac{1}{\sqrt{2\pi}}\int_c^\infty e^{-\frac{1}{2}y^2}\,dy$ is equal to the size of the critical region. Hence to maximize the power with respect to k is equivalent to maximizing

$$\psi(k) = \frac{E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k-1) - c\sqrt{2(k-1)}}{\sigma'}$$

with respect to k.

Under the alternative $F(\Delta, k)$ we obtain

$$E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k-1) = k(N-1)\sum_{i=1}^{i=k} p_i^2 + k - N - k + 1 = 4(N-1)\left(\Delta - \frac{1}{k}\right)^2$$

Hence

$$\psi(k) = \frac{4(N-1)\left(\Delta - \frac{1}{k}\right)^2 - c\sqrt{2(k-1)}}{\sigma'}.$$

We choose Δ so that this maximum power is exactly $\frac{1}{2}$, that is, so that $\psi(k) = 0$ for that k which maximizes $\psi(k)$. Denote this value of Δ by Δ_N and let k_N be the value of k which maximizes $\psi(k)$. The differential-quotient of the numerator of $\psi(k)$ with respect to k is then equal to 0 for $k = k_N$. Hence

$$8(N-1)\left(\Delta_N - \frac{1}{k_N}\right)\frac{1}{k_N^2} = \frac{c}{\sqrt{2(k_N-1)}}. \tag{10}$$

Furthermore since $\psi(k_N) = 0$ we have

$$4(N-1)\left(\Delta_N - \frac{1}{k_N}\right)^2 = c\sqrt{2(k_N-1)}. \tag{11}$$

Solving equations (10) and (11) we obtain

$$\Delta_N = \frac{5}{k_N} - \frac{4}{k_N^2} \tag{12}$$

and

$$\sqrt[5]{\frac{k_N^8}{(k_N-1)^5}} = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$$

or since $k_N > 3$,

$$k_N < 4 \sqrt[5]{\frac{2(N-1)^2}{c^2}} < k_N + 1.$$

Hence

$$\text{(13)} \quad \text{either } k_N = \left[4 \sqrt[5]{\frac{2(N-1)^2}{c^2}}\right] \quad \text{or} \quad k_N = \left[4 \sqrt[5]{\frac{2(N-1)^2}{c^2}}\right] + 1,$$

is the value of k for which the power with respect to $F(\Delta_N, k)$ becomes a maximum. We have merely to show that $\psi''(k)$ is negative for $k = k_N$.

Using the fact that $\psi(k_N) = \psi'(k_N) = 0$ we obtain

$$\sigma'\psi''(k_N) = \frac{-16(N-1)}{k_N^3}\Delta_N + \frac{24(N-1)}{k_N^4} + \frac{c}{(\sqrt{2(k_N-1)})^3}.$$

Substituting for Δ_N the right hand side of (12) we obtain on account of (10)

$$\sigma'\psi''(k_N) = \frac{-56(N-1)}{k_N^4} + \frac{64(N-1)}{k_N^5} + \frac{8(N-1)}{2(k-1)}\left(\frac{4}{k_N^3} - \frac{4}{k_N^4}\right).$$

Using $2(k-1) > k$ we obtain

$$\psi''(k_N) < \frac{1}{k^4\sigma'}\left(-24(N-1) + \frac{32}{k}(N-1)\right)$$

which is negative. σ' can be shown to be of order $k_N^{\frac{1}{2}}$; $\psi''(k_N)$ is, therefore, of order $\frac{N}{k_N^{4+\frac{1}{2}}} = 0\left(\frac{1}{N^{\frac{4}{5}}}\right)$. The maximum is, therefore, rather flat for large values of N.

We shall now show that if k is large enough to assume χ'^2 to be normally distributed then $F(\Delta, k)$ is the alternative which gives the smallest power compared with all alternatives in the class $C(\Delta)$ provided the power for the alternative $F(\Delta, k)$ equals $\frac{1}{2}$.

We know that $E\left(\sum_{i=1}^{i=k} x_i^2\right)$ is smallest for $F(\Delta, k)$ Since the power with respect to $F(\Delta, k)$ equals $\frac{1}{2}$ we have

$$E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k - 1 - c\sqrt{2(k-1)}) = 0$$

Thus the lower limit of the integral in (9) becomes negative for every other alternative and the power will be larger than $\frac{1}{2}$.

The power with respect to $F(\Delta_N, k_N)$ is equal to $\frac{1}{2}$, hence if we choose $k = k_N$ the power of the test will be $\geqslant \frac{1}{2}$ for all alternatives in the class $C(\Delta_N)$ On the other hand if we choose $k \neq k_N$ then there will be at least one alternative in

---

[3] Cantelli's formula and its proof are given by Fréchet in his book *Recherches Théoriques Modernes sur la Théorie de Probabilités*, Paris (1937), pp. 123–126

$C(\Delta_N)$ for which the power is $<\frac{1}{2}$. (For instance $F(\Delta_N, k)$ is such an alternative.)

The above statements have been derived under the assumption that χ'^2 is normally distributed. Hence if the distribution of χ'^2 were exactly normal $k_N = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$ would be a best k and for this k_N and $\Delta_N = \frac{5}{k_N} - \frac{4}{k_N^2}$ the greatest lower bound of the power in the class $C(\Delta_N)$ would be exactly $\frac{1}{2}$. Since the distribution of χ'^2 approaches the normal distribution with $k \to \infty$ the sequence $\{k_N\}$ is best in the limit and Theorem 1 stated in the introduction is proved.

For the purposes of practical applications, it is not enough to know that $\{k_N\}$ is best in the limit. We have to know for what values of N k_N can be considered practically as a best k, i.e. for what values of N the quantity $\epsilon(k_N, N)$ defined in the introduction is sufficiently small. The quantity $\epsilon(k_N, N)$ is certainly small if for the number of class intervals k_N the distribution of χ'^2 is near to normal and if the power with respect to at least one alternative of the class $C(\Delta_N)$ is smaller than $\frac{1}{2}$ also in the case when the number of class intervals is too small to assume a normal distribution for χ'^2.

We shall in the following assume that for $k > 13$ the normal distribution is a sufficiently good approximation. Actually we need not assume a normal distribution but only that the probability is close to $\frac{1}{2}$ that the statistic will exceed its mean value.

Cantelli[3] gave the following formula. Let M_r be the rth moment of a distribution about x_0. Let d be any arbitrary positive number. Let $P(|x - x_0| \leqslant d)$ be the probability that $|x - x_0| \leqslant d$ then the following inequalities hold:

$$\text{If}\quad \frac{M_r}{d^r} \leqslant \frac{M_{2r}}{d^{2r}} \quad\text{then}\quad P(|x - x_0| \leqslant d) \geqslant 1 - \frac{M_r}{d^r}.$$

$$\text{If}\quad \frac{M_r}{d^r} \geqslant \frac{M_{2r}}{d^{2r}} \quad\text{then}\quad P(|x - x_0| \leqslant d) \geqslant 1 - \frac{M_{2r} - M_r^2}{(d^r - M_r)^2 + M_{2r} - M_r^2}.$$

Since χ'^2 can only take positive values we have

$$\text{(14)}\quad \text{If}\quad \frac{E(\chi'^2)}{c_k} \leqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2} \quad\text{then}\quad P(\chi'^2 \leqslant c_k) \geqslant 1 - \frac{E(\chi'^2)}{c_k}.$$

$$\text{(15)}\quad \text{If}\quad \frac{E(\chi'^2)}{c_k} \geqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2}$$

$$\text{then}\quad P(\chi'^2 \leqslant c_k) \geqslant 1 - \frac{\sigma^2_{\chi'^2}}{(c_k - E(\chi'^2))^2 + \sigma^2_{\chi'^2}}.$$

Where c_k is determined so that $P(\chi^2 \geqslant c_k)$ equals the size of the critical region if the null hypothesis is true and the number of class intervals equals k. c_k can be obtained from a table of the chi-square distribution.

For $F(\Delta_N, k)$ we obtain with $\Delta_N' = \frac{5}{k_N} - \frac{4}{k_N^2} - \frac{1}{k}$ from (3) and (4)

$$E(\chi'^2) = (k - 1) + 4(N - 1)\Delta_N'^2,$$
$$\sigma^2_{\chi'^2} = 2(k - 1) + 8\Delta_N'^2(k + 2N - 4) - 32(2N - 1)\Delta_N'^4.$$

By numerically calculating $E(\chi'^2)$ and $\sigma_{\chi'^2}$ for $N = 450$ and a 5% level of significance, for $N = 300$ and a 1% level of significance, and for $k = 13, 12 \cdots \left[\frac{1}{\Delta_N}\right] + 1$ it can be shown that for these values of N and k

$$\text{(16)} \qquad \frac{E(\chi'^2)}{c_k} \geqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2}.$$

Hence we have to use (15). From (16) it follows that $c_k > E(\chi'^2)$. If $P(\chi'^2 \leqslant c_k \leqslant \frac{1}{2}$ we obtain on account of (15) and (16)

$$\frac{\sigma^2_{\chi'^2}}{(c_k - E(\chi'^2))^2 + \sigma^2_{\chi'^2}} \geqslant \frac{1}{2}, \qquad \sigma_{\chi'^2} + E(\chi'^2) \geqslant c_k.$$

Numerical calculation shows that for the values of N and k and the significance levels considered

$$\text{(17)} \qquad \sigma_{\chi'^2} + E(\chi'^2) < c_k.$$

It can then be shown that for $N \geqslant 450$ and $N \geqslant 300$ respectively $N\Delta_N'$ decreases with N. A simple argument then shows that (16) and (17) are also true for all values $N \geqslant 450$ and $N \geqslant 300$ respectively. Hence the power with respect to $F(\Delta_N, k)$ is $<\frac{1}{2}$ for these values of N. Thus we see· For $N \geqslant 450$ if the 5% level is used, and for $N \geqslant 300$ if the 1% level is used, the value $k_N = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$ can be considered for practical purposes as a best k The value c is determined so that $\frac{1}{\sqrt{2\pi}}\int_c^\infty e^{-\frac{1}{2}t^2}\,dt$ is equal to the size of the critical region.

LIMITED TYPE OF PRIMARY PROBABILITY DISTRIBUTION APPLIED TO ANNUAL MAXIMUM FLOOD FLOWS

By Bradford F. Kimball

Port Washington, N. Y.

1. Theoretical statement of problem. There is no doubt that Gumbel's recent paper "The Return Period of Flood Flows" [1] has supplied an admirably simple technique for engineers to use in approximating the trend of *return periods* of annual maximum flood flows for purposes of extrapolation. This treatment is scientifically of great interest because it introduces for the first time into a subject already treated at considerable length by engineers, the theory of the probability distribution of maximum values as developed by Fisher and Tippett, von Mises, and others.[1] However, certain further observations should be made concerning the approach used by Gumbel.

Let x represent the measure of daily stream flow having a probability distribution $w(x)$. Let the probability distribution of the associated annual maximum stream flows be denoted by $V(x)$ with

$$W(x) = \int_0^x V(s)\, ds, \tag{1}$$

denoting probability that annual maxima be less than or equal to x. The *return period* $T(x)$ of an annual maximum flow of measure x is then defined by

$$T(x) = \frac{1}{1 - W(x)}. \tag{2}$$

In this paper the probability distribution $w(x)$ will be called the *primary* probability distribution associated with the probability distribution of maximum values $V(x)$ and its *cumulative* distribution $W(x)$.

Gumbel argues that for the type of primary probability distribution that might reasonably be expected to apply, $W(x)$ will be of the type introduced by R. A. Fisher:

$$W(x) = \exp\left[-\exp - \alpha(x - u)\right]. \tag{3}$$

It is further implied that a primary probability distribution involving an upper limit would lead to a probability distribution of maximum values of the type

$$W_1(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \cdot e^{-(u/x)^k}, \tag{4}$$

for which moments of order k or higher do not exist. The inference is then drawn that a primary probability distribution leading to such a cumulative distribution of maximum values would seem to be less likely to be the correct

[1] See references at end of Gumbel's paper, loc. cit.

one than one leading to the distribution (3). To this argument we do not object; but we question the implied conclusion that *hence the use of a limited type of primary distribution is to be disallowed.*

If the primary probability distribution be of the *limited* Galton type

$$w(x) = K \exp\left(-\tfrac{1}{2}u^2\right), \tag{5}$$

where K is a constant and

$$u = k[b - \log (a - x)], \qquad 0 \leqq x \leqq a, \tag{6}$$

it can be shown that the limiting form of the cumulative distribution of maxima of n values takes the same type form (3) where x is replaced by u. This can be seen by observing that the transformed variate u becomes infinite as x approaches a, and hence has infinite range to the right, which places (5) in the category of distributions which are known to lead to cumulative distribution of maximum values of form (3). More explicitly, considering $w(x)$ as a finite distribution in x, if one traces the reasoning as set forth in von Mises' derivation [2] of the limiting distribution (3), one finds that since the cumulative primary probability $\int_0^x w(s)\, ds$ does *not* have a non-vanishing derivative of finite order at $x = a$, that what von Mises terms *the case of a limited distribution* does *not* apply, while the argument for a cumulative distribution of maxima of form (3) *does* carry through, in spite of the fact that x has limited range to the right. This fact was not mentioned by Gumbel

One is thus led to the conclusion that there is no logical exclusion of the assumption of a primary probability distribution of the form (5).

One might well argue for a first approximation of the actual primary probability distribution of stream flows—using any regular time interval such as a day or an hour—of the form (5). Differentiating u with respect to x, one obtains

$$k\, dx = (a - x)\, du, \tag{7}$$

which means that to a constant probability increment Δu there corresponds a maximum increment Δx in measure of stream flow equal to $(a/k)\Delta u$ when x is at the lower limit zero. This corresponding increment in stream flow decreases linearly to zero as x approaches its upper bound a, imposed because of the existence of a finite watershed.

2. Technique of fitting probability distribution of maximum values in case primary probability distribution is of the limited type (5)–(6). Write the cumulative maximum distribution (3) in the form:

$$\begin{gathered} W(x) = \exp(-\exp -y), \qquad y = \alpha(u(x) - u_1), \\ u(x) = k[b - \log (a - x)], \qquad 0 \leqq x \leqq a. \end{gathered} \tag{8}$$

Now it is known that for the distribution

$$dW = e^{-e^{-y}} e^{-y} \, dy, \tag{9}$$

the mean value and standard deviation of y are given by

$$\begin{aligned} \bar{y} &= .577215 \text{ (Euler's constant } C) \\ \sigma^2(y) &= \pi^2/6. \end{aligned} \tag{10}$$

Hence

$$\bar{y} = \alpha[\bar{u}(x) - u_1] = \alpha k[(b - u_1/k) - \bar{L}] = C$$

where $\bar{L}$ denotes the mean value of $\log (a - x)$, with x representing the observed *maximum flood flows.* Also

$$\sigma(y) = \alpha k \, \sigma(L) = \pi/\sqrt{6}$$

where $\sigma(L)$ denotes the standard deviation of $\log (a - x)$. Hence

$$\alpha k = (\pi/\sqrt{6})/\sigma(L), \qquad b - u_1/k = C/\alpha k + \bar{L}, \tag{11}$$

and y is determined as a function of x by the relation

$$y = \alpha k[(b - u_1/k) - \log (a - x)]. \tag{12}$$

It is interesting to observe that it has not been necessary to determine the constants k and b of the primary probability distribution. Only the upper bound a and observed flood flows are used in this process. From the relation (12) the theoretical curve in terms of x may easily be computed from tables relating y to W (See Gumbel, loc. cit., Table II, page 173).

The difficulty of determining what the upper bound a should be in a specific case is a practical one and does not concern the objective theoretical problem of choosing the *type* of curve which most nearly describes the behavior of annual maximum flood flows. The point to be made in this paper is that the use of what seems to be a reasonable value of a, *will* materially alter forecasts of future annual flood flows relative to forecasts made on the assumption that such an upper limit may be neglected. It is also ventured that the resulting theoretical probability distribution of maxima will in general give a better fit to the series of observed floods than one based on the latter premise. Techniques for determination of upper bound a will not be discussed in this paper.

3. Examples. In order to demonstrate the point in question the two methods have been applied to a 57 year record of the annual flood flows of the Tennessee River at Chattanooga for the years 1875 to 1931.[2]

---

[2] The author has already used this series in a previous article [3] and for this reason has found it convenient to use it here.

TABLE I

Series of observed annual flood flows

(Tennessee River at Chattanooga, 1875–1931)

(1) Observed Flood x	(2) Ratio to Mean	(3) Per cent of Time	(4) Return Period, $T(x)$
85.9	.412	0.88	1.007
108	.518	2.63	1.027
123	.590	4.39	1.043
			....
310	1.487	95.61	22.8
349	1.674	97.37	38.0
361	1.731	99.12	114.

In Table I, col. (1) is shown the incomplete series of observed annual floods in units of 1,000 c.f.s. arranged in order of magnitude. The complete series may be referred to in *Water-Supply Paper 771* entitled "Floods in the United States," *U. S. Geological Survey*, 1936, p. 401 The mean annual maximum flood of this series is 208.56. The ratio of each annual maximum to the mean is shown in Col. (2). In Col. (4) is shown the observed return period which is taken here as the harmonic mean between what has been called the *exceedance interval* and the *recurrence interval* (see Gumbel, loc. cit., Table I, p. 167). Thinking of the 57 year record as a span of 57 years, the above procedure is equivalent to taking the observed probability $W(x)$ that a given annual flood will not be exceeded as the mid-point of the part of this time-span covered by the observed flood in question. Thus the lowest flood-peak 85,900 c.f s. corresponds to the span from zero to 1.754 per cent of the whole time-span, and hence $W(x)$ is taken at the mid-point, -0.877 per cent. Similarly the greatest flood, 361,000 c.f.s. corresponds to interval from 98.246 to 100 per cent and is taken at 99.12 per cent. These arithmetic means correspond to harmonic means of the "recurrence" and "exceedance" intervals referred to above. This is the procedure which Hazen [4] originally followed.

Data from Cols. (1) and (4) of this table determined position of dots on Fig. 1. Data from Cols (2) and (3) gave the points indicated by dots on Fig. 2, with $1 - W(x)$ recorded on the chart rather than $W(x)$.

The two theoretical distributions fitted to these annual flood maxima will be referred to as distributions A and B.

Distribution A. In this case the limited type of primary probability distribution (5) – (6) is assumed From previous studies of this data series made by the author [3], an upper limit of annual floods of some 609,000 c f s. was found to be reasonable, and for purposes of this example the same upper limit will be assumed for the primary probability distribution. Thus the transformation (6) becomes:

$$u = k[b - \log (609 - x)],$$

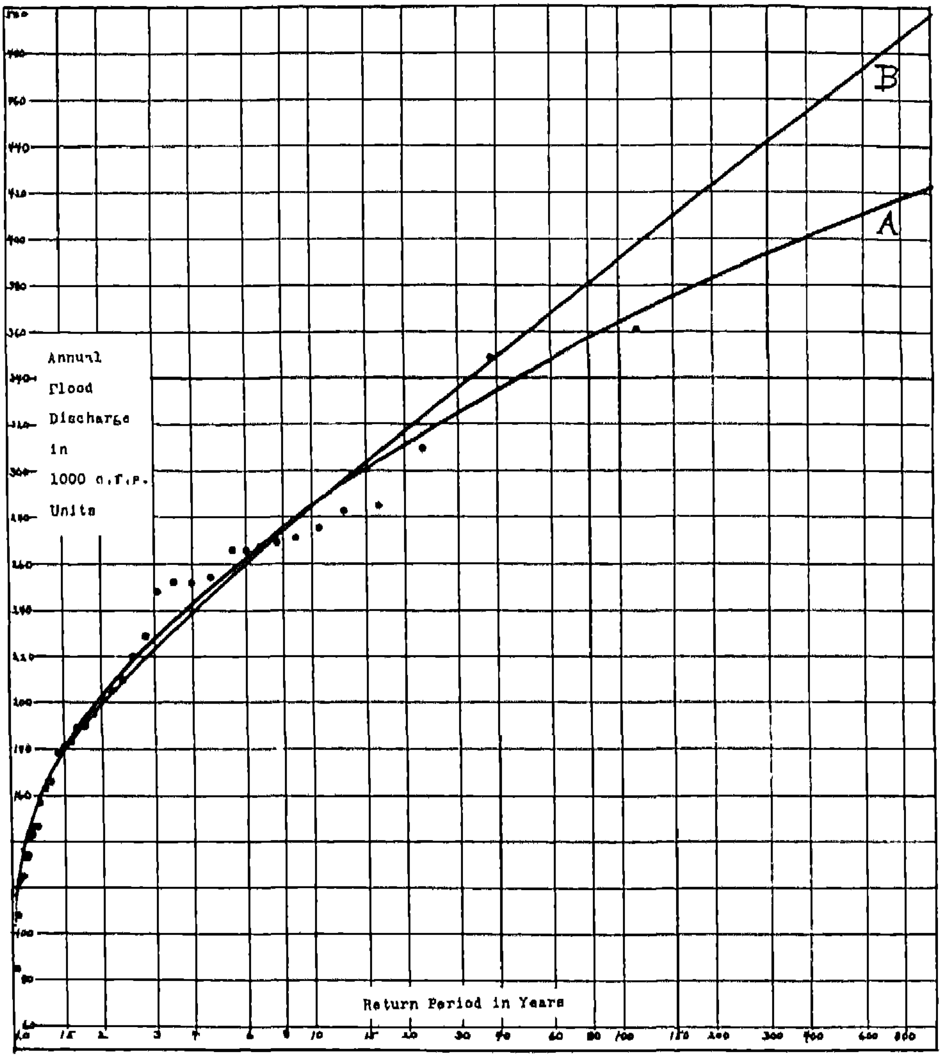


FIG. 1. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—return periods plotted against annual flood discharges on semi-logarithmic chart.

where the logarithm to base 10 can be used without loss of generality since the constant k will absorb the conversion factor. The mean value of the logarithm, and its standard deviation come to

$$\bar{L} = 2.59772, \qquad \sigma(L) = .06576$$

The constants of the transformation (12) are thus determined by

$$\alpha k = (\pi/\sqrt{6})/(.06576), \qquad b - u_1/k = C/(\alpha k) + 2.59772$$

Thus

$$1/(\alpha k) = .05127, \qquad b - u_1/k = 2.6273$$

and solving (12) for log (609 − x),

$$\log (609 - x) = 2.6273 - (.05127)\, y \tag{13}$$

Using a table for the known relations between y, $W(x)$, and $T(x)$ for the Fisher-Tippett distribution of maximum values similar to Table II of Gumbel's article

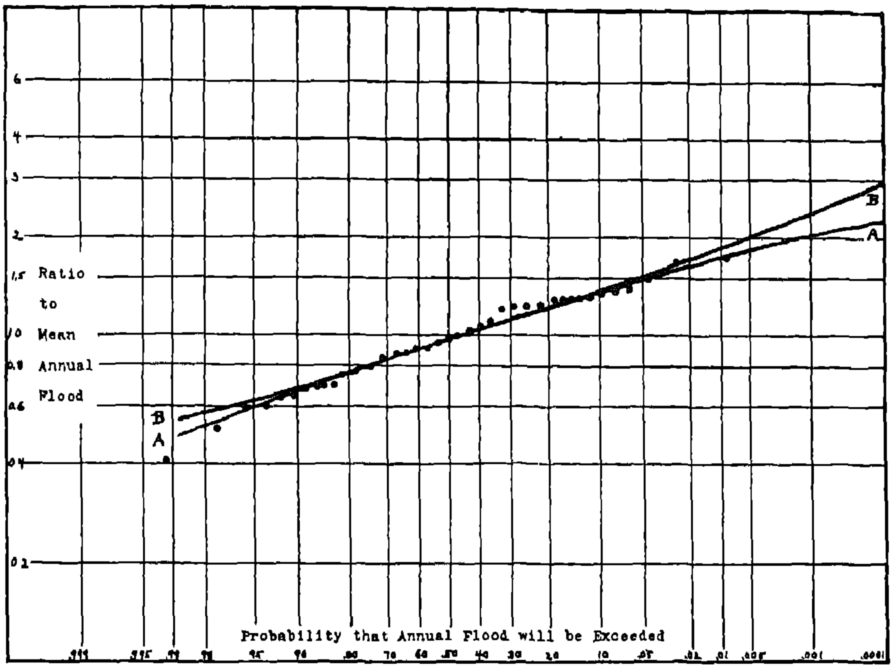


FIG. 2. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—Data plotted on logarithmic probability chart designed by Hazen, Whipple and Fuller

(loc. cit.) the corresponding values x of the annual floods are easily determined. Thus a theoretical relation between x and $W(x)$ is set up. This is indicated as Curve A on the two charts exhibited here.

Distribution B. The primary probability distribution in this case is taken as unlimited to the right, and in general is assumed to have the character of an exponentially decreasing function of the measure of stream flow x (see Gumbel, loc. cit.). The parameter y of the distribution of annual maxima is given directly by

$$y = \alpha(x - x_1)$$

and

$$1/\alpha = (\sqrt{6}/\pi) \text{ (stand. dev. of annual floods)} = (.77970)\ (58.26) = 45.425$$

$$x_1 = \text{(mean annual flood)} - C/\alpha = 208.6 - (.57722)\ (45.425) = 182.4$$

Hence

$$x = 182.4 - (45.425)\, y \tag{14}$$

and using the table of corresponding values of y, $W(x)$ and $T(x)$ for the Fisher-Tippett distribution referred to above, a theoretical relation between x and $W(x)$ is easily set up. This is plotted as Curve B on the accompanying charts.

4. Discussion of examples. In Fig. 1 it is to be noted that if theoretical curves are continued to the right to give readings for a return period of 1,000 years, the divergence of Curve A from Curve B is large enough to be of significance, numerically. Visual inspection does not indicate which curve is the better fit to the observation points.

In Fig. 2 the curves are plotted on "logarithmic probability" graph paper. This paper was designed by Hazen and Fuller [4] specifically for the purpose of plotting annual maxima of stream-flows. A significant divergence in trend is to be noted at the right hand end.

These charts indicate that the use of an upper limit may materially affect extrapolation of fitted theoretical curves, for purposes of estimating floods with a return period, say of 1,000 years.

If the trends of observed floods in Gumbel's recent paper in the *Transactions of the American Geophysical Union* [5] are examined, it will be observed that in the case of the Connecticut, Mississippi and Rhone rivers, there is a decided tendency for the curve of observed floods to turn downwards, away from the theoretical curves, which correspond to Curve B exhibited in Figure 1. In the case of the Tennessee, Cumberland and Columbia rivers the tendency is not decisive, while in the case of the Rhine river at Basel (Switzerland) the tendency of the observed curve is upwards rather than downwards. As the writer has observed elsewhere [6], this last data series seems to be rather unique in character and is possibly the result of a watershed greatly influenced by all year around snow deposits. Possibly a radically different primary probability distribution should be used in this case.

5. Conclusion. The writer has demonstrated in this paper that in fitting a theoretical probability distribution of maximum values to annual maxima of stream flows, the use of an upper bound for measures of stream flow by assumption of a primary probability distribution of the type (5)–(6)

(1) is not inconsistent with the use of the Fisher-Tippett distribution of maxima,

(2) has a reasonable logical basis from the point of view of the hydrologist,

(3) may materially affect the estimation of return periods when extrapolation is involved, relative to results obtained when no upper bound is assumed.

It has not been within the scope of this paper to discuss techniques for determining such an upper bound, nor to apply the theory to enough data series to draw conclusions concerning goodness of fit.

REFERENCES

[1] E. J. Gumbel, "The return period of flood flows," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 163–190

[2] Richard von Mises, "La distribution de la plus grande de n valeurs," *Revue Mathematique de l'Union Interbalkanique*, Vol 1, Athens (1936)

[3] Bradford F. Kimball, "Probability-distribution-curve for flood-control studies," *Trans Am Geographical Union*, 1938, pp. 466–475

[4] Allen Hazen, *Flood-flows*, John Wiley & Sons, New York, 1930.

[5] E J. Gumbel, "Probability-interpretation of the observed return-periods of floods," *Trans Am. Geophysical Union*, 1941, pp 836–850.

[6] Bradford F Kimball, *Discussion* of paper entitled "Statistical control curves for flood discharges" by E. J Gumbel, *Trans Am. Geophysical Union*, 1942

LINEAR RESTRICTIONS ON CHI-SQUARE

By Franklin E. Satterthwaite

University of Iowa

Chi-square is a statistic widely used in statistical analysis. It is usually of the form,

$$
\begin{aligned}
\chi^2 &= \sum_1^n \chi_i^2 \\
&= \sum_1^n \left(\frac{x_i - m_i}{\sigma_i}\right)^2,
\end{aligned}
\tag{1}
$$

where the x_i's are independent normally distributed variables drawn from populations with respective means and standard deviations, m_i and σ_i. In practical problems the independence of the x_i's is often modified by placing restrictions on the χ_i's in order to estimate the m_i's or σ_i's. It is well known that if m such restrictions which are linear and homogeneous (also algebraically independent) are placed on the χ_i's, then the resulting chi-square, (1), is distributed according to the chi-square distribution with $n - m$ degrees of freedom. The purpose of this paper is to study the case where the restrictions are not necessarily homogeneous.

1. Geometrical development. The χ_i's of equation (1) may be considered as co-ordinates in an n-dimensional space. Equation (1) represents a sphere in such a space with its center at the origin and with radius, χ. We should like to determine the distribution of χ^2. First, since the χ_i's are independent, we may form their joint distribution,[1]

$$
\begin{aligned}
F(\chi_1, \chi_2, \cdots \chi_n)\, dV &= K\Pi_j e^{-\frac{1}{2}\chi_j^2}\, d\chi_j \\
&= Ke^{-\frac{1}{2}\Sigma\chi_j^2}\, d\chi_1\, d\chi_2 \cdots d\chi_n \\
&= Ke^{-\frac{1}{2}\chi^2}\, dV.
\end{aligned}
\tag{2}
$$

We may change the variable in (2) to χ^2 if we can determine dV. Since the n-dimensional sphere represented by equation (1) has a volume proportional to χ^n, we may write

$$
\begin{aligned}
dV &= Kd(\chi^2)^{\frac{1}{2}n} \\
&= K(\chi^2)^{\frac{1}{2}n-1}\, d\chi^2.
\end{aligned}
$$

Substituting this value in the distribution (2) we obtain for the distribution of chi-square,

$$F(\chi^2)\, d\chi^2 = K(\chi^2)^{\frac{1}{2}n-1} e^{-\frac{1}{2}\chi^2}\, d\chi^2,$$

which is the usual form of the chi-square distribution for n degrees of freedom.

---

[1] The letter K will be used throughout as a constant, not necessarily the same constant from equation to equation.

We shall next restrict the values of χ_i by means of a condition,

$$a_{11}\chi_1 + a_{12}\chi_2 + \cdots + a_{1n}\chi_n = \rho_1, \qquad \Sigma a_{1j}^2 = 1, \tag{3}$$

where ρ_1 is a constant. This restriction represents a hyper-plane in our n-dimensional space at a distance ρ_1 from the origin. The intersection of this hyper-plane with our sphere (1) is an $(n - 1)$-dimensional sphere with radius

$$\chi' = (\chi^2 - \rho_1^2)^{\frac{1}{2}}.$$

The differential of the volume of this sphere is

$$dV = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1}\, d\chi^2.$$

Substituting this in the distribution (2) we obtain the distribution of chi-square subject to the single linear restriction, (3). Thus

$$F(\chi^2)\, d\chi^2 = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1} e^{-\frac{1}{2}\chi^2}\, d\chi^2,$$

or more conveniently,

$$F(\chi^2 - \rho_1^2)\, d(\chi^2 - \rho_1^2) = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1} e^{-\frac{1}{2}(\chi^2-\rho_1^2)}\, d(\chi^2 - \rho_1^2)$$

The argument may be readily extended to include additional linear restrictions of the form,

$$\begin{array}{ll} a_{21}\chi_1 + a_{22}\chi_2 + \cdots + a_{2n}\chi_n = \rho_2, & \Sigma a_{2j}^2 = 1, \\ \cdots & \\ a_{m1}\chi_1 + a_{m2}\chi_2 + \cdots + a_{mn}\chi_n = \rho_m, & \Sigma a_{mj}^2 = 1. \end{array} \tag{4}$$

For convenience we shall assume that the restrictions form an orthogonal set[2] so that

$$\Sigma_j a_{ij} a_{kj} = 0, \qquad i \neq k.$$

The hyper-plane represented by equation (4) is at a distance, ρ_2, from the origin. Since (4) is orthogonal to (3), it is also at a distance, ρ_2, from the center of the $(n - 1)$-dimensional sphere obtained on applying the first restriction. Therefore the intersection of this hyper-plane with the $(n - 1)$-dimensional sphere will give an $(n - 2)$-dimensional sphere of radius

$$\chi'' = (\chi^2 - \rho_1^2 - \rho_2^2)^{\frac{1}{2}}.$$

Similarly, if we consider all m restrictions, we obtain an $(n - m)$-dimensional sphere with radius

$$\chi^{(m)} = (\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}}.$$

[2] Any set of linear restrictions which are algebraically independent and consistent may be replaced by an orthogonal set. Thus if (4) were not orthogonal to (3), we could replace (4) by (4) $- k$(3) where k is determined by the condition

$$\Sigma a_{1j}(a_{2j} - k a_{1j}) = 0$$

or

$$\Sigma a_{1j} a_{2j} = k \Sigma a_{1j}^2$$

The differential of the volume of this sphere will be

$$dV = K(\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}(n-m)-1}\, d(\chi^2 - \Sigma\rho_j^2).$$

Substituting this in (2) we see that

$$(\chi^{(m)})^2 = \chi^2 - \sum \rho_j^2$$

is distributed as is chi-square with $n - m$ degrees of freedom.

2. Alternate analytic development. It is perhaps desirable that we present an analytic proof of the foregoing theorem. Therefore we shall first regard the ρ_j's as variables and shall determine the joint distribution of χ^2 and the ρ_j's. We may then pass to the distribution of those values of χ^2 which correspond to assigned values of the ρ_j's. Note that the χ_i's are considered to be statistically independent.

The characteristic function of the joint distribution of χ^2 and the ρ_j's is known to be[3]

$$\phi(t, t_1, \cdots t_m) = \frac{e^{-Q/2(1-2it)}}{(1-2it)^{\frac{1}{2}n}},$$

where

$$Q = \sum_{i,j,k} a_{ik} a_{jk} t_i t_j$$
$$= \sum t_i^2, \qquad \text{since } \sum a_{ik} a_{jk} = \delta_{ij}.$$

Applying the Fourier transform, we obtain the joint distribution of χ^2 and the ρ_j's:

$$F(\chi^2, \rho_1, \cdots \rho_n) = K \int \cdots \int \frac{e^{Q'}}{(1-2it)^{\frac{1}{2}n}}\, dt_m \cdots dt_1\, dt,$$

where

$$Q' = -it\chi^2 - \Sigma i t_j \rho_j - \{\Sigma t_j^2/2(1-2it)\}$$
$$= -it\chi^2 - \frac{\Sigma[t_j + i\rho_j(1-2it)]^2}{2(1-2it)} - \tfrac{1}{2}(1-2it)\Sigma\rho_j^2.$$

Performing the integration with respect to $t_1, \cdots t_m$, we have,

$$F = Ke^{-\frac{1}{2}\Sigma\rho_j^2} \int \frac{e^{-it\chi^2}}{(1-2it)^{\frac{1}{2}n}} (1-2it)^{\frac{1}{2}m} e^{it\Sigma\rho_j^2}\, dt,$$

and finally,

$$F = K(\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}(n-m)-1} e^{-\frac{1}{2}\chi^2}.$$

[3] See A. T. Craig, "A certain mean value problem in statistics," *Bull. Amer. Math. Soc.*, Vol. 42 (1936), p. 671.

In our problem we want the distribution of χ^2 (or more conveniently, of $\chi^2 - \Sigma\rho_j^2$) when the ρ_j's take on fixed values. To obtain this we substitute fixed values, $\hat{\rho}_j$'s, into the joint distribution and divide by the marginal total,

$$\int F(\chi^2, \hat{\rho}_1, \hat{\rho}_2 \cdots \hat{\rho}_m)\, d\chi^2 = K\Gamma[\tfrac{1}{2}(n-m)]2^{\frac{1}{2}(n-m)}\, e^{-\frac{1}{2}\Sigma\hat{\rho}_j^2}.$$

This gives us the distribution function,

$$F(\chi^2 - \Sigma\hat{\rho}) = \frac{1}{2\Gamma[\frac{1}{2}(n-m)]}[\tfrac{1}{2}(\chi^2 - \Sigma\hat{\rho}_j^2)]^{\frac{1}{2}(n-m)-1}\, e^{-\frac{1}{2}(\chi^2 - \Sigma\hat{\rho}_j^2)},$$

which is a chi-square distribution with $n - m$ degrees of freedom.

3. Application. As an example of the use of linear restrictions on chi-square we shall now examine the effect on the chi-square test of goodness of fit if the moments of a sample are not corrected for grouping errors in fitting a frequency curve.

The parameters of the fitted frequency distribution, $f(x)$, are determined from the equations,

$$N\int x^k f(x)\, dx = \Sigma\, x_j^k \theta_j, \qquad k = 0, 1, 2, \cdots, \tag{5}$$

where x_j is the mid-point of the j^{th} group and θ_j the corresponding observed frequency. Next a set of expected frequencies,

$$\hat{\theta}_j = \int_{\alpha_j}^{\alpha_j+1} Nf(x)\, dx, \qquad \alpha_j = (x_{j-1} + x_j)/2,$$

is determined by taking partial areas of the fitted frequency distribution. The expected frequency is used to transform the actual frequency into a statistic with mean zero and unit variance by the equation,

$$\chi_j = (\theta_j - \hat{\theta}_j)/\hat{\theta}^{\frac{1}{2}}.$$

Equations (5) may now be rearranged into the form of linear restrictions on the χ_j. Thus

$$\Sigma\, x_j^k \hat{\theta}_j^{\frac{1}{2}} \chi_j = \rho_k' \tag{6}$$

where the ρ_k' have the values,

$$\begin{aligned} \rho_k' &= \sum x_j^k \theta_j - \Sigma\, x_j^k \hat{\theta}_j \\ &= N\int x^k f(x)\, dx - \Sigma\, x_j^k \hat{\theta}_j \\ &\neq 0 \text{ in general} \end{aligned}$$

To make our example more specific, let us fit a normal distribution to a sample of 1000 items with mean zero and unit variance. Let the grouping be about the midpoints,

$$x_j: \qquad -3, \quad -2, \quad -1, \quad 0, \quad 1, \quad 2, \quad 3.$$

The expected frequencies in each group are

$$\hat{\theta}_j. \qquad 6, \quad 61, \quad 242, \quad 382, \quad 242, \quad 61, \quad 6.$$

The variance of these expected frequencies is 1.080 as contrasted with 1.000 for the sample. The linear restrictions, (6), now take the forms,

$$(7) \quad 2.4\chi_{-3} + 7.8\chi_{-2} + 15.6\chi_{-1} + 19.5\chi_0 + 15.6\chi_1 + 7.8\chi_2 + 2.4\chi_3 = 0$$

$$(8) \quad -7.2\chi_{-3} - 15.6\chi_{-2} - 15.6\chi_{-1} + 0 + 15.6\chi_1 + 15.6\chi_2 + 7.2\chi_3 = 0$$

$$(9) \quad 21.6\chi_{-3} + 31.2\chi_{-2} + 15.6\chi_{-1} + 0 + 15.6\chi_1 + 31.2\chi_2 + 21.6\chi_3 = -80.$$

Because of the symmetry of the normal distribution, restriction (8) is orthogonal to (7) and (9). Therefore the only orthogonalization necessary is to replace (9) by an equivalent restriction which is orthogonal to (7). This can be done by subtracting 1.080 times (7) from (9) which gives

$$(10) \quad 19.0\chi_{-3} + 22.8\chi_{-2} - 1.2\chi_{-1} - 21.1\chi_0 - 1.2\chi_1 + 22.8\chi_2 + 19.0\chi_3 = -80$$

If these restrictions are each divided by the square root of the sum of the squares of the coefficients of the χ_j, they will be the normal orthogonal set required by the development. The distances of these restrictive planes from the center of χ^2-sphere are

$$\rho_{(7)} = 0, \qquad \rho_{(8)} = 0, \qquad \rho_{(10)} = 1.7.$$

Thus if we test the goodness of fit of the normal distribution to this sample by calculating chi-square,

$$\chi^2 = \Sigma\, \chi_j^2 = \Sigma \frac{(\theta_j - \hat{\theta}_j)^2}{\hat{\theta}_j},$$

we should subtract from χ^2 a correction of

$$\Sigma\, \rho_k^2 = 2.8$$

before judging the significance. This correction adjusts for the effect of the grouping error on the chi-square test.

In this example, chi-square has four degrees of freedom so that an error of 2.8 is large enough to affect our judgment of its significance. It can be shown that the correction is proportional to the size of the sample. Therefore, if our sample had contained only 100 items, the fit obtained by ignoring grouping effects would be almost as good as the fit when the sample moments were corrected for grouping. On the other hand, if the sample had 10,000 items, it

would be practically impossible to obtain a satisfactory fit without correcting for grouping errors.

4. Conclusion. The theory of the loss of degrees of freedom for chi-square when the underlying statistics are subject to linear restrictions does not require the restrictions to be homogeneous. For restrictions which are not homogeneous, a correction must be subtracted from chi-square equal to the square of the distance from the center of the sphere,

$$\chi^2 = \Sigma\, \chi_i^2 = 0$$

to the intersection of the restrictive planes. Non-homogeneous restrictions sometimes arise in practice because of the bias introduced by an approximation. An example is given from curve fitting.

SYSTEMS OF LINEAR EQUATIONS WITH COEFFICIENTS SUBJECT TO ERROR

By A. T. Lonseth

Iowa State College

1. Introduction. Various scientific problems lead to non-homogeneous systems of n linear equations in n unknowns, in which the $n^2 + n$ coefficients (including "absolute" terms) are subject to error. Such errors may be errors of observation, or errors introduced by rounding off decimal expansions. If the system has a non-vanishing determinant, the ordinary rules yield the solution. But the question arises: how may the possible errors in the coefficients affect the solutions? In particular, one would like to know how to exclude the fatal event that some malicious combination of errors might make the determinant zero. One would further like to have limitations on the solution-errors in terms of maximum coefficient-errors. Considering the coefficient-errors as random variables, one may also inquire as to the probability distributions of the solution-errors.

The principal result obtained in this paper is the Taylor's expansion of the error in any unknown, considered as a function of the $n(n + 1)$ errors in the coefficients. An upper bound is obtained for each term of this series, and the sum of these upper bounds (when convergent) is expressed in closed form. Thus are obtained not only approximations to the maximum error, but an actual upper limit. Convergence of the power series is established for sufficiently small coefficient-errors; "sufficient smallness" is specified in terms of a simple criterion, which simultaneously provides a sufficient condition for the non-vanishing of a determinant with elements subject to error.

These results were obtained before I learned that work had already been done on the problem. The earliest seems to be that of F. R. Moulton [2] in 1913; he found the first order approximation (6) for $n = 3$, and discussed the geometrical reasons for sensitivity. Much later I. M. H. Etherington [1], evidently unaware of Moulton's paper, found the expression for the total error of a determinant whose elements may be in error, and applied this to the present problem. He thus found limits for the first and second order errors, in a rather different form from mine. The probabilistic considerations of section 5 were suggested by Etherington's article. L. B. Tuckerman [3] recently discussed the question of estimating computational errors incurred in the course of solution. He considered only errors of first order.

My original procedure was to compute the terms of the Taylor's series as successive differentials of the unknown, from Cramer's formula. This soon becomes laborious, and I found only the first two terms. The linear matrix equation (4) was then kindly suggested to me by R. Oldenburger. Here (4) is solved by iteration, resulting in a simple recursion formula for successive terms of the Taylor's series.

2. Formal matrix solution. Let the system of equations be

$$\sum_{j=1}^{n} a_{ij} x_j = c_i \qquad i = 1, 2, \cdots, n. \tag{1}$$

In terms of the matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

system (1) can be written

$$\mathbf{AX} = \mathbf{C}. \tag{2}$$

Supposing that not all c's vanish, and that A, the determinant of $\mathbf{A}$, does not vanish, there is a unique solution $\mathbf{X}$. But the a's and c's, and consequently the x's, are subject to error: let the true value of a_{ij} be $a_{ij} + \alpha_{ij}$; of c_i, $c_i + \gamma_i$; and of the resulting x_j, $x_j + \xi_j$. We must actually deal with the system

$$(\mathbf{A} + \mathbf{a})(\mathbf{X} + \mathbf{x}) = \mathbf{C} + \mathbf{c}, \tag{3}$$

where we have written

$$\mathbf{a} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Expanding (3) and using (2), we find for the error-matrix $\mathbf{x}$

$$\mathbf{x} = \mathbf{m} + \mathbf{nX} + \mathbf{nx}, \tag{4}$$

with $\mathbf{m} = \mathbf{A}^{-1}\mathbf{c}$, $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{a}$; $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$. We solve (4) formally for $\mathbf{x}$ by iteration. Thus

$$\mathbf{x} = \mathbf{m} + \mathbf{nX} + \mathbf{n}(\mathbf{m} + \mathbf{nX}) + \mathbf{n}^2\mathbf{x}, \text{ etc.}:$$

and there results the infinite expansion

$$\mathbf{x} = \sum_{k=1}^{\infty} \mathbf{x}^{(k)}: \mathbf{x}^{(1)} = \mathbf{m} + \mathbf{nX}; \qquad \mathbf{x}^{(k)} = \mathbf{nx}^{(k-1)}, \qquad k > 1. \tag{5}$$

In section 4 convergence of (5) will be established for sufficiently small $|\alpha_{ij}|$.

3. The elements of $\mathbf{x}^{(k)}$. It is necessary to consider closely the individual elements of $\mathbf{x}^{(k)}$. Writing

$$\mathbf{x}^{(k)} = \begin{pmatrix} \xi_1^{(k)} \\ \vdots \\ \xi_n^{(k)} \end{pmatrix},$$

we note from (5) that

$$\xi_j = \sum_{k=1}^{\infty} \xi_j^{(k)} ;$$

this is precisely the Taylor's series for the error in x_j : each $\xi_j^{(k)}$ is a homogeneous polynomial of degree k in the α's and γ's. Writing A_{ij} for the cofactor of a_{ij} in A,

$$\begin{aligned}\mathbf{x}^{(1)} &= \mathbf{m} + \mathbf{n}\mathbf{X} = \mathbf{A}^{-1}(\mathbf{c} - \mathbf{a}\mathbf{X}) \\ &= \begin{pmatrix} \frac{A_{11}}{A} & \cdots & \frac{A_{n1}}{A} \\ \vdots & & \vdots \\ \frac{A_{1n}}{A} & \cdots & \frac{A_{nn}}{A} \end{pmatrix} \left\{ \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} - \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\} \\ &= \begin{pmatrix} \frac{A_{11}}{A} & \cdots & \frac{A_{n1}}{A} \\ \vdots & & \vdots \\ \frac{A_{1n}}{A} & \cdots & \frac{A_{nn}}{A} \end{pmatrix} \begin{pmatrix} \gamma_1 - \alpha_{11} x_1 - \cdots - \alpha_{1n} x_n \\ \vdots \\ \gamma_n - \alpha_{n1} x_1 - \cdots - \alpha_{nn} x_n \end{pmatrix},\end{aligned}$$

whence (summing hereafter from 1 to n on Greek-letter subscripts)

$$\xi_j^{(1)} = \frac{1}{A} \left\{ \sum_{\mu} \gamma_\mu A_{\mu j} - x_1 \sum_{\mu} \alpha_{\mu 1} A_{\mu j} - \cdots - x_n \sum_{\mu} \alpha_{\mu n} A_{\mu j} \right\}. \tag{6}$$

From (5), if $k > 1$,

$$\begin{aligned}\mathbf{x}^{(k)} &= \mathbf{n}\mathbf{x}^{(k-1)} = -\mathbf{A}^{-1}\mathbf{a}\mathbf{x}^{(k-1)} \\ &= \begin{pmatrix} -\frac{1}{A}\Sigma\alpha_{\mu 1}A_{\mu 1} & \cdots & -\frac{1}{A}\Sigma\alpha_{\mu n}A_{\mu 1} \\ \vdots & & \vdots \\ -\frac{1}{A}\Sigma\alpha_{\mu 1}A_{\mu n} & \cdots & -\frac{1}{A}\Sigma\alpha_{\mu n}A_{\mu n} \end{pmatrix} \begin{pmatrix} \xi_1^{(k-1)} \\ \vdots \\ \xi_n^{(k-1)} \end{pmatrix},\end{aligned}$$

so that

$$\xi_j^{(k)} = -\frac{1}{A} \sum_{\nu} \xi_\nu^{(k-1)} \sum_{\mu} \alpha_{\mu\nu} A_{\mu j}, \qquad k > 1. \tag{7}$$

The sums $\Sigma\gamma_\mu A_{\mu j}$, $\Sigma\alpha_{\mu i}A_{\mu j}$ have obvious interpretations as determinants.

4 Bounds and convergence of the series. Assuming $|\alpha_{ij}|, |\gamma_i| \leqq \delta$ and taking absolute values in (6),

$$|\xi_j^{(1)}| \leqq \frac{\delta}{|A|}(1 + \sum_{\mu} |x_\mu|)(\sum_{\mu} |A_{\mu j}|). \tag{8}$$

It will be observed that equality can be attained for a particular choice of α's and γ's as $\pm\delta$: the bound for first-order errors is best possible. But it is not in general possible by a single choice of α's and γ's to obtain equality for all j.

Similarly from (7)

$$|\xi_j^{(k)}| \leqq \frac{\delta}{|A|}\Big(\sum_\mu |\xi_\mu^{(k-1)}|\Big)\Big(\sum_\mu |A_{\mu j}|\Big), \qquad k > 1,$$

whence by induction

$$|\xi_j^{(k)}| \leqq \left(\frac{\delta}{|A|}\right)^k \Big(1 + \sum_\mu |x_\mu|\Big)\Big(\sum_\nu \sum_\mu |A_{\mu\nu}|\Big)^{k-1}\Big(\sum_\mu |A_{\mu j}|\Big) \tag{9}$$

Summing on k,

$$\sum_{k=1}^m |\xi_j^{(k)}| \leqq \frac{\delta}{|A|}\Big(1 + \sum_\mu |x_\mu|\Big)\Big(\sum_\mu |A_{\mu j}|\Big)\left(\sum_{k=1}^m \rho^{k-1}\right),$$

with

$$\rho = \frac{\delta}{|A|}\sum_\nu \sum_\mu |A_{\mu\nu}|.$$

If $\rho < 1$, we can let $m \to \infty$:

$$|\xi_j| \leqq \frac{\delta}{|A|}\Big(1 + \sum_\mu |x_\mu|\Big)\Big(\sum_\mu |A_{\mu j}|\Big)/(1 - \rho). \tag{10}$$

Observing that the γ's occur linearly in (6) and (7), we conclude that (5) converges if

$$|\alpha_{ij}| \leqq \delta < |A|/\Big(\sum_\nu \sum_\mu |A_{\mu\nu}|\Big). \tag{11}$$

It follows that the determinant of the system (3) cannot vanish if (11) holds. This is rather remarkable, in that $\delta\Sigma\Sigma\ |A_{\mu\nu}|$ is merely the maximum first-order term in the error of that determinant ([1], p. 108); the effect of higher order terms (i e., of any but first-order minors) in producing a zero determinant can be wholly ignored.

From the remark after (8), it appears that equality in (9) and (10) cannot generally be attained.

If (10) is written $|\xi_j| \leqq B/(1 - \rho)$, it is easily seen that the remainder after the hth approximation does not exceed $\rho^h B/(1 - \rho)$

5. Probability distributions. We now consider some consequences of the following assumptions: the α's and γ's are identical, independent random variables, bounded by a δ satisfying (11), and distributed symmetrically about zero. (It would be reasonable to assume further that they possess a frequency function, which is nowhere concave upward) Writing $\mathfrak{E}(x)$ for "expectation of the random variable x," we have

$$\mathfrak{E}(\alpha_{ij}) = \mathfrak{E}(\gamma_i) = 0, \qquad \mathfrak{E}(\alpha_{ij}^2) = \mathfrak{E}(\gamma_i^2) = \sigma^2 < \delta^2.$$

On account of independence and symmetry, the expectation of any power-product of α's and γ's containing an odd power must be zero. To first order, the mean a_j of the solution-error ξ_j is approximated by

(12) $$a_j^{(1)} = \mathcal{E}(\xi_j^{(1)}) = 0;$$

and the standard deviation S_j by

(13) $$S_j^{(1)} = \sqrt{\mathcal{E}\{(\xi_j^{(1)})^2\}} \frac{\sigma}{|A|} \{(1 + \sum_\mu x_\mu^2)(\sum_\mu A_{\mu j}^2)\}^{\frac{1}{2}}.$$

The second approximation to a_j is also easily obtained:

(14) $$a_j^{(2)} = \mathcal{E}(\xi_j^{(2)}) = \frac{\sigma^2}{A^2}(\sum_\nu \sum_\mu x_\nu A_{\mu\nu} A_{\mu j}).$$

Both (13) and (14) were given by Etherington [1], though in a less symmetric form. Higher approximations, as he remarks, involve complicated summations; but if they should ever be required, the machinery exists in (6) and (7) for their systematic computation. As to the errors in using (13) for the standard deviation S_j and (14) for the mean, we know only that

$$a_j = a_j^{(2)} + o(\delta^4), \qquad S_j^2 = (S_j^{(1)})^2 + o(\delta^4).$$

Etherington ([1], p. 111) considers the important special case of "rounding off" decimal expressions. Each a and c is supposed correct in the qth decimal place, the $(q + 1)$th figure being "forced," i.e., increased by one when the $(q + 2)$th figure is dropped, if the $(q + 2)$th is 5, 6, 7, 8, or 9. Assuming constant frequency 10^{-q} in the interval $(-\frac{1}{2}10^{-q}, \frac{1}{2}10^{-q})$, we may use (13) and (14) with $\sigma^2 = 10^{-2q}/12$.

Errors of observation are often assumed to be normally distributed. There is nothing against such an assumption with regard to the γ's, but the α's must not make (3) singular, and must accordingly be suitably bounded, e.g. by (11).

6. Conclusion. The formulas and bounds of this paper involve only these quantities: the determinant A, its first order minors, and the solutions of (1). They can be found in the course of solving (1) by orthodox methods.

Inequality (10) definitely limits the maximum solution-errors, in terms of the maximum coefficient-error δ, provided δ satisfies (11). But it may be that (8), either alone or in conjunction with the second-order bound from (9), will give a better approximation.

The ratio $\Sigma\Sigma \mid A_{\mu\nu} \mid / \mid A \mid$ may be taken as a "measure of sensitivity" of (1) to error.

The fundamental formulas (6) and (7) are capable of solving other problems than those studied here. For example, it may happen that only certain elements (such as those of a single column) are in error, in which case better inequalities can be found. Or the α's and γ's may not be independently and identically distributed.

REFERENCES

[1] I. M. H. ETHERINGTON, "On errors in determinants," *Proc Edinburgh Math. Soc*, Ser 2, Vol 3 (1932), pp. 107–117.

[2] F. R MOULTON, "On the solutions of linear equations having small determinants," *Amer. Math. Monthly*, Vol 20 (1913), pp 242-249.

[3] L. B TUCKERMAN, "On the mathematically significant figures in the solution of simultaneous linear equations," *Annals of Math Stat.*, Vol 12 (1941), pp 307–316

[4] P G. HOEL, "The errors involved in evaluating correlation determinants," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 58–65

ON MUTUALLY FAVORABLE EVENTS

By Kai-Lai Chung

Tsing Hua University, Kunming, China

Introduction. For a set of arbitrary events, E. J. Gumbel, M. Fréchet and the author[1] have recently obtained inequalities between sums of certain probability functions. One of the results of the author is the following:

Let $E_1, \cdots, E_n$ be n arbitrary events and let $p_m(\nu_1, \cdots, \nu_k)$ denote the probability of the occurrence of at least m events out of the k events $E_{\nu_1}, \cdots, E_{\nu_k}$. Then, for $k = 1, \cdots, n-1$ and $1 \leqq m \leqq k$ we have

$$\binom{n-m}{k-m} \Sigma p_m(\nu_1, \cdots, \nu_{k+1}) \leqq \binom{n-m}{k-m+1} \Sigma p_m(\nu_1, \cdots, \nu_k),$$

where the summations extend respectively to all combinations of $k+1$ and k indices out of the n indices $1, \cdots, n$.

In course of proof of the above inequalities it appears that similar inequalities between products instead of sums can be obtained under certain assumptions regarding the nature of interdependence of the events. We shall first study the nature of such assumptions, and then proceed to the proof of the said inequalities (Theorems 1 and 2). It may be noted that the inductive method used here serves equally well for the proof of the inequalities cited above, though somewhat longer, but apparently our former method is not applicable here.

That events satisfying our assumptions actually exist, is shown by an application to the elementary theory of numbers. The author feels incompetent to discuss other possible fields of application.

1. Let a set of events be given

$$E_1, E_2, \cdots, E_n, \cdots$$

and let E'_i denote the event non-E_i. Let $p(i)$ denote the probability of the occurrence of E_i, $p(i')$ that of the occurrence of E'_i. For convenience we assume that for any i $p_i(1 - p_i) \neq 0$; events with the exceptional probabilities 0 or 1 may evidently be left out of account.

Let $p(\nu_1 \cdots \nu_k)$ denote the probability of the occurrence of the conjunction $E_{\nu_1} \cdots E_{\nu_k}$ and let $p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)$ denote the probability of the occurrence of $E_{\nu_1} \cdots E_{\nu_k}$, on the hypothesis that $E_{\mu_1} \cdots E_{\mu_h}$ have occurred. The μ's or ν's may be accented.

Definition 1: *If $p(\nu_1, \nu_2) > p(\nu_2)$, we say that the occurrence of the event E_{ν_1} is favorable to the occurrence of the event E_{ν_2}, or simply that E_{ν_1} is favorable to E_{ν_2}.*

[1] "On the probability of the occurrence of at least m events among n arbitrary events," *Annals of Math. Stat.* Vol 12 (1941), pp. 328–338.

If $p(\nu_1, \nu_2) = p(\nu_2)$, we say that E_{ν_1} is indifferent to E_{ν_2}. If $p(\nu_1, \nu_2) < p(\nu_2)$, we say that E_{ν_1} is unfavorable to E_{ν_2}.

Thus the relations "favorableness," "indifference," and "unfavorableness" are mutually exclusive and together exhaustive. We state the following immediate consequences:

(i) Reflexity: An event is favorable to itself; in fact, $p(\nu, \nu) = 1 > p(\nu)$.

(ii) Symmetry: If E_1 is favorable (indifferent, unfavorable) to E_2, then E_2 is favorable (indifferent, unfavorable) to E_1. In fact, we have

$$p(1)p(1, 2) = p(12) = p(2)p(2, 1),$$

$$\frac{p(1, 2)}{p(2)} = \frac{p(2, 1)}{p(1)}.$$

Thus $p(1, 2) \gtreqless p(2)$ is equivalent to $p(2, 1) \gtreqless p(1)$.

In particular, if E_1 is indifferent to E_2, then so is E_2 to E_1. They are then usually said to be independent of each other.

(iii) If E_1 is favorable (indifferent, unfavorable) to E_2, then E_1' is unfavorable (indifferent, favorable) to E_2. For, we have

$$p(1)p(1, 2) + p(1')p(1', 2) = p(12) + p(1'2) = p(2),$$

whence

$$p(1')p(1', 2) = p(2) - p(1)p(1, 2).$$

On the other hand,

$$p(1')p(2) = [1 - p(1)]p(2) = p(2) - p(1)p(2).$$

Since by assumption $p(1')p(2) \neq 0$, we have

$$\frac{p(1', 2)}{p(2)} = \frac{p(2) - p(1)p(1, 2)}{p(2) - p(1)p(2)}.$$

Thus

$$p(1', 2) \gtreqless p(2) \text{ according as } p(1, 2) \gtreqless p(2).$$

For the sake of brevity we introduce the following symbolic notation:

$$E_1/E_2 = \begin{cases} 1, \text{ if } E_1 \text{ is favorable to } E_2 \\ 0, \text{ if } E_1 \text{ is indifferent to } E_2 \\ -1, \text{ if } E_1 \text{ is unfavorable to } E_2. \end{cases}$$

Then by (ii) and (iii) we have

$$E_1/E_2 = E_2/E_1,$$

$$E_1'/E_2 = E_2/E_1' = E_1/E_2' = E_2'/E_1 = -(E_1/E_2),$$

$$E_1'/E_2' = E_2'/E_1' = E_1/E_2,$$

analogous to the rules of signs in the multiplication of integers.

(iv) Non-transitivity: If E_1 is favorable to E_2, and E_2 is favorable to E_3, it does not necessarily follow that E_1 is favorable to E_3; in fact, it may happen that E_1 is unfavorable to E_3. For instance, imagine 11 identical balls in a bag marked respectively with the numbers

$$-11,\ -10,\ -3,\ -2,\ -1,\ 2,\ 4,\ 6,\ 11,\ 13,\ 16.$$

Let a ball be drawn at random. Let

$$E_1 = \text{(the event of the number on the ball being positive)}$$

$$E_2 = \text{(the event of the number on the ball being even)}$$

$$E_3 = \text{(the event of the number on the ball being of 1 digit)}$$

We have

$$p(1, 2) = \tfrac{4}{6} > \tfrac{6}{11} = p(2),$$
$$p(2, 3) = \tfrac{4}{6} > \tfrac{6}{11} = p(3),$$
$$p(1, 3) = \tfrac{1}{2} < \tfrac{6}{11} = p(3).$$

(v) It may happen that $E_1/E_3 = 1$, $E_2/E_3 = 1$, but $E_1E_2/E_3 = -1$. In the example above,

$$p(2, 1) = \tfrac{4}{6} > \tfrac{6}{11} = p(1),$$
$$p(3', 1) = \tfrac{3}{5} > \tfrac{6}{11} = p(1),$$
$$p(23', 1) = \tfrac{1}{2} < \tfrac{6}{11} = p(1).$$

(vi) It may happen that $E_1/E_2 = 1$, $E_1/E_3 = 1$, but $E_1/E_2E_3 = -1$. Example:

$$p(1, 2) = \tfrac{4}{6} > \tfrac{6}{11} = p(2),$$
$$p(1, 3') = \tfrac{1}{2} > \tfrac{5}{11} = p(3'),$$
$$p(1, 23') = \tfrac{1}{6} < \tfrac{2}{11} = p(23').$$

(vii) It may happen that $E_1/E_3 = 1$, $E_2/E_3 = 1$, but the disjunction $(E_1 + E_2)/E_3 = -1$. For, by (v) we know that there exist events E_1', E_2', E_3' such that

$$E_1'/E_3' = 1, \qquad E_2'/E_3' = 1, \qquad E_1'E_2'/E_3' = -1.$$

Hence by (iii) there exist events E_1, E_2, E_3 such that

$$E_1/E_3 = 1, \qquad E_2/E_3 = 1, \qquad (E_1'E_2')'/E_3 = -1.$$

But $(E_1'E_2')' = E_1 + E_2$. Thus the last relation is $(E_1 + E_2)/E_3 = -1$.

(viii) It may happen that $E_1/E_2 = 1$, $E_1/E_3 = 1$, but $E_1/(E_2 + E_3) = -1$. This follows from (vi) as (vii) follows from (v).

After all these negative results in (iv)–(viii), we see that we cannot expect to go far without making stronger assumptions regarding the nature of inter-

dependence between the events in the set. Firstly, in view of (iv), we shall restrict ourselves to consideration of a set of events in which each event is favorable to every other. Secondly, in view of (v), we shall only consider the case where the "favorableness," as defined above, shall be cumulative in its effect, that is to say, the more events favorable to a given event have been known to occur, the more probable this given event shall be esteemed. We formulate these two conditions in mathematical terms, as follows:

DEFINITION 2: *A set of events $E_1, \cdots, E_n, \cdots$ is said to be strongly mutually favorable (in the first sense) if, for every integer h and every set of distinct indices (positive integers) $\mu_1, \cdots, \mu_h$ and ν we have*

$$p(\mu_1 \cdots \mu_h, \nu) > p(\mu_1 \cdots \mu_{h-1}, \nu).$$

This definition requires that there exist no implication relation between any event and any conjunction of events in the set; in particular, that the events are all distinct. It would be more convenient to consider the relation "favorable or indifferent to." This will be done later on. The present definitions have the advantage of being logically clear cut and also that of yielding unambiguous inequalities.

From Definition 2 we deduce the following consequences:

(1) If the set $(\mu_1^*, \cdots, \mu_j^*)$ is a sub-set of $(\mu_1, \cdots, \mu_h)$, we have

$$p(\mu_1 \cdots \mu_h, \nu) > p(\mu_1^* \cdots \mu_j^*, \nu).$$

(2) For any positive integer k and any two sets $(\nu_1, \cdots, \nu_k)$ and $(\mu_1, \cdots, \mu_h)$ where all the indices are distinct, we have

$$p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k) > p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k).$$

More generally, we have as in (1),

$$p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k) > p(\mu_1^* \cdots \mu_j^*, \nu_1 \cdots \nu_k).$$

PROOF: We have only to prove the first inequality. For $k = 1$ this is the assumption in Definition 2. Suppose that the inequality holds for $k - 1$, we shall prove that it holds for k, too.

$$\frac{p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k)} = \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1} \nu_1 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_h, \nu_1)p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_{h-1}, \nu_1)p(\mu_1 \cdots \mu_{h-1} \nu_1, \nu_2 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_h, \nu_1)}{p(\mu_1 \cdots \mu_{h-1}, \nu_1)} \frac{p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1} \nu_1, \nu_2 \cdots \nu_k)}$$

$$> \frac{p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1} \nu_1, \nu_2 \cdots \nu_k)} > 1.$$

Observe that none of the denominators vanish by our original assumption and by Definition 2.

Therefore we see that when the failure in (v) is remedied by our definition, the failure in (vi) is automatically remedied too.

2. THEOREM 1: *Let $n > 1$ and let $E_1, \cdots, E_n, \cdots$ be a set of strongly mutually favorable events (in the first sense). Then we have, for $k = 1, \cdots, n-1$,*

$$\prod_{\nu_1, \cdots, \nu_{k+1}} [p(\nu_1 \cdots \nu_{k+1})]^{\binom{n-1}{k}^{-1}} > \prod_{\nu_1, \cdots, \nu_k} [p(\nu_1 \cdots \nu_k)]^{\binom{n-1}{k-1}^{-1}}$$

where the products extend respectively to all combinations of $k + 1$ and k distinct indices out of the indices $1, \cdots, n$.

PROOF. We may assume that the indices are written so that

$$1 \leqq \nu_1 < \nu_2 < \cdots < \nu_{k+1} \leqq n.$$

Taking logarithms, we have

$$\binom{n-1}{k-1} \sum_{\nu_1, \cdots, \nu_{k+1}} \log p(\nu_1 \cdots \nu_{k+1}) > \binom{n-1}{k} \sum_{\nu_1, \cdots, \nu_k} \log p(\nu_1 \cdots \nu_k).$$

Substituting from the obvious formula

$$p(\nu_1 \cdots \nu_k) = p(\nu_1)p(\nu_1, \nu_2)p(\nu_1\nu_2, \nu_3) \cdots p(\nu_1 \cdots \nu_{k-1}, \nu_k),$$

and writing $\log p(\cdots) = q(\cdots)$, the inequality becomes

$$\begin{aligned} &\binom{n-1}{k-1} \Sigma[q(\nu_1) + q(\nu_1, \nu_2) + \cdots + q(\nu_1 \cdots \nu_k, \nu_{k+1})] \\ &\qquad > \binom{n-1}{k} \Sigma[q(\nu_1) + q(\nu_1, \nu_2) + \cdots + q(\nu_1 \cdots \nu_{k-1}, \nu_k)]. \end{aligned} \tag{1}$$

Immediately we observe that the number of terms of the form $q(\nu_1 \cdots \nu_s, \mu)(0 \leqq s \leqq \mu - 1)$ with a fixed μ after the comma in the bracket is the same on both sides, since

$$\binom{n-1}{k-1}\binom{n-1}{k} = \binom{n-1}{k}\binom{n-1}{k-1}. \tag{2}$$

Let the sums of such q's on the left and right of (1) be $\sigma^{(1)} = \sigma^{(1)}(\mu)$ and $\sigma^{(2)} = \sigma^{(2)}(\mu)$ respectively. To prove our theorem it is sufficient to prove that $\sigma^{(1)}(\mu) \geqq \sigma^{(2)}(\mu)$ for every μ and $\sigma^{(1)}(\mu) > \sigma^{(2)}(\mu)$ for at least one μ.

Now the terms in $\sigma^{(1)}$ (or $\sigma^{(2)}$) fall into classes according to the number s of the μ_i's before the comma in the bracket. Let those terms having s μ_i's before the comma belong to the s-th class. It is evident that the number of terms of the s-th class in $\sigma^{(1)}(\mu)$ is equal to

$$\binom{n-1}{k-1}\binom{\mu-1}{s}\binom{n-\mu}{k-s}$$

for $s = 0, 1, \cdots, \mu - 1$; where we make the convention that

$$\binom{0}{0} = 1, \qquad \binom{a}{b} = 0 \quad \text{if } a < b \text{ or if } b < 0.$$

Thus for a fixed μ, when the terms in $\sigma^{(1)}(\mu)$ are classified in the above manner, its total number of terms may be written as the following sum, in which vanishing terms may occur:

$$\binom{n-1}{k-1}\binom{n-1}{k} = \binom{n-1}{k-1}\left\{\binom{\mu-1}{\mu-1}\binom{n-\mu}{k-\mu+1} + \binom{\mu-1}{\mu-2}\binom{n-\mu}{k-\mu+2} + \cdots + \binom{\mu-1}{s}\binom{n-\mu}{k-s} + \cdots + \binom{\mu-1}{0}\binom{n-\mu}{k}\right\}.$$

Similarly the total number of terms in $\sigma^{(2)}(\mu)$ may be written as the following sum:

$$\binom{n-1}{k}\binom{n-1}{k-1} = \binom{n-1}{k}\left\{\binom{\mu-1}{\mu-1}\binom{n-\mu}{k-\mu} + \binom{\mu-1}{\mu-2}\binom{n-\mu}{k-\mu+1} + \cdots + \binom{\mu-1}{s}\binom{n-\mu}{k-s-1} + \cdots + \binom{\mu-1}{0}\binom{n-\mu}{k-1}\right\}.$$

LEMMA 1: *For $0 \leqq s \leqq k$, we have, taking account of our conventions about the binomial coefficients,*

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} > \binom{n-1}{k}\binom{n-\mu}{k-s-1} \quad \textit{for} \quad s > (\mu-1)k/n; \tag{3}$$

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} \leqq \binom{n-1}{k}\binom{n-\mu}{k-s-1} \quad \textit{for} \quad s \leqq (\mu-1)k/n \tag{4}$$

PROOF: Suppose $s \geqq k - n + \mu$, then

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} \gtreqless \binom{n-1}{k}\binom{n-\mu}{k-s-1}$$

according as

$$\frac{k}{n-k} \gtreqless \frac{k-s}{n-\mu-k+s+1},$$

i.e. according as

$$s \gtreqless (\mu-1)k/n.$$

But, since $k < n$ and $\mu \leqq n$, we have

$$n - k - k/n + 1 > (n-k)\mu/n$$
$$(\mu-1)k/n > k - n + \mu - 1$$

so that

$$(\mu - 1)k/n + 1 \geqq k - n + \mu.$$

Therefore if $s > (\mu - 1)k/n$, then $s \geqq (\mu - 1)k/n + 1 \geqq k - n + \mu$, and (3) holds.

Again, if $k - n + \mu \leqq s \leqq (\mu - 1)k/n$, then (4) holds; while if $s < k - n + \mu$, then the left-hand side of (4) vanishes while the right-hand side is non-negative, thus (4) holds for $s \leqq (\mu - 1)k/n$. The lemma is proved.

If we put $(s = 0, 1, \cdots, k)$

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} - \binom{n-1}{k}\binom{n-\mu}{k-s-1} = d_s,$$

then by Lemma 1,

$$d_s \gtreqless 0 \quad \text{according as} \quad s \gtreqless (\mu - 1)k/n.$$

This means that although the total number of terms of the form $p(\mu_1 \cdots \mu_s, \mu)$ is the same on both sides of (1), the left-hand side is more abundant in terms with larger s while the right-hand side is more abundant in terms with smaller s. Now we have

$$q(\mu_1 \cdots \mu_i, \mu) > q(\mu_1^* \cdots \mu_j^*, \mu)$$

if $i > j$ and if $(\mu_1^* \cdots \mu_j^*)$ is a subset of $(\mu_1 \cdots \mu_i)$. Hence it is natural to suppose that the left-hand side must be greater because it is more abundant in terms of larger values. Unfortunately even if $i > j$, the last inequality is in general not true if the set $(\mu_1^* \cdots \mu_j^*)$ is not a sub-set of $(\mu_1 \cdots \mu_i)$. Therefore we cannot as yet conclude that $\sigma^{(1)} \geqq \sigma^{(2)}$.

To prove that actually we have $\sigma^{(1)} \geqq \sigma^{(2)}$, we make the following "process of compensation":

We have, by (2) and the definition of d_s, the following equality:

$$\binom{\mu-1}{0}d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{\mu-1}d_{\mu-1} = 0.$$

where $d_j = 0$ if $j > k$. Thus

$$d_s \leqq 0 \quad \text{for} \quad s \leqq k(\mu - 1)/n,$$

$$d_s \geqq 0 \quad \text{for} \quad s > k(\mu - 1)/n.$$

Hence

$$\binom{\mu-1}{0}d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{\mu-1}d_{\mu-1} \leqq 0 \tag{5}$$

for $\quad l = 0, 1, \cdots, \mu - 1.$

For the fixed μ, let

$$\rho_l^{(1)} = \binom{n-1}{k-1}\left\{\binom{n-\mu}{k}q_\mu + \binom{n-\mu}{k-1}\sum_{\mu_1<\mu} q(\mu_1, \mu) + \cdots + \binom{n-\mu}{k-l}\sum_{\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu)\right\}$$

$$\rho_l^{(2)} = \binom{n-1}{k}\left\{\binom{n-\mu}{k-1}q_\mu + \binom{n-\mu}{k-2}\sum_{\mu_1<\mu} q(\mu_1, \mu) + \cdots + \binom{n-\mu}{k-l-1}\sum_{\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu)\right\}$$

so that

$$\rho_{\mu-1}^{(1)} = \sigma^{(1)}(\mu), \qquad \rho_{\mu-1}^{(2)} = \sigma^{(2)}(\mu).$$

For $\mu = 1$, $l = 0$, we have

$$\sigma^{(1)}(1) = \rho_0^{(1)} = \rho_0^{(2)} = \sigma^{(2)}(1).$$

LEMMA 2: *For $\mu > 1$ and $0 \leqq l < \mu - 1$, we have*

$$\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu) < \frac{l+1}{\mu-l-1}\sum_{1\leqq\mu_1<\cdots<\mu_{l+1}<\mu} q(\mu_1 \cdots \mu_{l+1}, \mu).$$

PROOF: We have, for any $\nu < \mu$, $\nu \neq \mu_i$ $(i = 1, \cdots, l)$

$$q(\mu_1 \cdots \mu_l\nu, \mu) > q(\mu_1 \cdots \mu_l, \mu).$$

Summing with respect to all such ν's,

$$\sum_\nu q(\mu_1 \cdots \mu_l\nu, \mu) > (\mu - l - 1)q(\mu_1 \cdots \mu_l, \nu).$$

Summing with respect to all $1 \leqq \mu_1 < \cdots < \mu_l < \mu$,

$$\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu}\sum_\nu q(\mu_1 \cdots \mu_l\nu, \mu) = (l+1)\sum_{1\leqq\mu_1<\cdots<\mu_{l+1}<\mu} q(\mu_1 \cdots \mu_{l+1}, \mu) > (\mu - l - 1)\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu).$$

The lemma is proved.

Now we use induction to prove that for $\mu > 1$ and $l = 1, \cdots, \mu - 1$

$$\rho_l^{(1)} - \rho_l^{(2)} > \frac{d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{l}d_l}{\binom{\mu-1}{l}} \times \sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu) \geqq 0.$$

This inequality holds for $l = 1$ by Lemma 2. Assume that it holds for l, $(l < \mu - 1)$. Then we have, by (5) and the fact that each $q < 0$,

$$\rho_{l+1}^{(1)} - \rho_{l+1}^{(2)} = \rho_l^{(1)} - \rho_l^{(2)} + d_{l+1} \sum_{1 \leqq \mu_1 < \cdots < \mu_{l+1} < \mu} q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$> \frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l}{\binom{\mu-1}{l}} \sum q(\mu_1 \cdots \mu_l, \mu) + d_{l+1} \sum q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$\geqq \left(\frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l}{\binom{\mu-1}{l}} \frac{l+1}{\mu - l - 1} + d_{l+1} \right) \sum q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$= \frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l + \binom{\mu-1}{l+1} d_{l+1}}{\binom{\mu-1}{l+1}} \sum q(\mu_1 \cdots \mu_{l+1}, \mu) \geqq 0.$$

Therefore, for $\mu > 1$, we have

$$\sigma^{(1)}(\mu) - \sigma^{(2)}(\mu) = \rho_{\mu-1}^{(1)} - \rho_{\mu-1}^{(2)} > 0.$$

Since $n > 1$ and $1 \leqq \mu \leqq n$, there exists a $\mu > 1$. Hence

$$\sum_{\mu=1}^{n} \sigma^{(1)}(\mu) > \sum_{\mu=1}^{n} \sigma^{(2)}(\mu)$$

which is equivalent to the inequality (1).

3. Our next step will be to obtain a generalization of Theorem 1. Consider a derived event defined by a disjunction of a (finite) number of events in the set, as follows.

$$E_{\nu_1} + E_{\nu_2} + \cdots + E_{\nu_m}.$$

We call such a disjunction a disjunction of the m-th order.

DEFINITION 3: *A set of events is said to be strongly mutually favorable in the second sense if for every positive integer m, the derived set of events consisting of all the disjunctions of the m-th order forms a strongly mutually favorable set of events (in the first sense).*

Let $D = D(m)$ denote in general a disjunction of the m-th order; let $p(D_1 \cdots D_h, D)$ denote the probability of the occurrence of the disjunction D, on the hypothesis that the conjunction of the h disjunctions $D_1 \cdots D_h$ has occurred. Then Definition 3 says that for any positive integer h and any set of distinct D's we have

$$p(D_1 \cdots D_h, D) > p(D_1 \cdots D_{h-1}, D).$$

Since a disjunction of the 1st order is an event E, we see that Definition 3 includes Definition 2.

Let $D_m(\nu_1, \cdots, \nu_k)$, $\nu_1 < \cdots < \nu_k$ denote the derived event

$$\prod_{\mu_1, \cdots, \mu_m} (E_{\mu_1} + \cdots + E_{\mu_m})$$

where the product (conjunction) extends to all combinations of m indices out of the indices $\nu_1, \cdots, \nu_k$. Let $p_m^*(\nu_1, \cdots, \nu_k)$ denote the probability of the occurrence of $D_m(\nu_1, \cdots, \nu_k)$. It is seen that $p_1^*(\nu_1, \cdots, \nu_k) = p(\nu_1 \cdots \nu_k)$ in our previous notation.

We merely state Theorem 2, whose proof is analogous to that of Theorem 1 but requires more cumbersome expressions.

THEOREM 2: *Let $n > k \geqq m \geqq 1$ and let $E_1, \cdots, E_n$ be a set of mutually strongly favorable events in the second sense. Then we have*

$$\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} [p_m^*(\nu_1, \cdots, \nu_{k+1})]^{\binom{n-m}{k-m+1}^{-1}} > \prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} [p_m^*(\nu_1, \cdots, \nu_k)]^{\binom{n-m}{k-m}^{-1}}.$$

To give an interpretation of $p_m^*(\nu_1, \cdots, \nu_k)$, we prove the symbolic equation between events.

$$D_m = \prod_{\nu_1 \leqq \mu_1 < \cdots < \mu_m \leqq \nu_k} (E_{\mu_1} + \cdots + E_{\mu_m}) = \sum_{\nu_1 \leqq \mu_1 < \cdots < \mu_{k-m+1} \leqq \nu_k} (E_{\mu_1} \cdots E_{\mu_{k-m+1}}) = C_{k-m+1},$$

where product means conjunction and sum means disjunction.

To prove this, we write for a general event E, $E = 1$ when E occurs, $E = 0$ when E does not occur. Now if $C_{k-m+1} = 0$, then at most $k - m$ events among the k given events occur, so that there exist m events such that $E_{\lambda_1} = 0$, $E_{\lambda_2} = 0$, $E_{\lambda_m} = 0$, thus

$$E_{\lambda_1} + E_{\lambda_2} + \cdots + E_{\lambda_m} = 0$$

Now the last disjunction is contained in D_m as a factor, therefore $D_m = 0$.

Conversely, if $D_m = 0$, at least one of its factors $= 0$, so that there exist m events, such that $E_{\lambda_1} = 0$, $E_{\lambda_2} = 0$, $\cdots$, $E_{\lambda_m} = 0$. Thus at most $k - m$ events out of the k given events occur and so by definition $C_{k-m+1} = 0$ Q.e.d.

From the above it immediately follows that

$$p_m^*(\nu_1, \cdots, \nu_k) = p_{k-m+1}(\nu_1, \cdots, \nu_k)$$

where $p_{k-m+1}(\nu_1, \cdots, \nu_k)$ is defined in the Introduction. Then Theorem 2 may be written as

$$\Pi[p_{k-m+2}(\nu_1, \cdots, \nu_{k+1})]^{\binom{n-m}{k-m+1}^{-1}} > \Pi[p_{k-m+1}(\nu_1, \cdots, \nu_k)]^{\binom{n-m}{k-m}^{-1}}$$

or again as

$$\Pi[w_{m-1}(\nu_1', \cdots, \nu_{k+1}')]^{\binom{n-m}{k-m+1}^{-1}} > \Pi[w_{m-1}(\nu_1', \cdots, \nu_k')]^{\binom{n-m}{k-m}^{-1}}$$

where $w_{m-1}(\nu_1', \cdots, \nu_k')$ denotes the probability of the occurrence of at most $m - 1$ events out of the k events $E'_{\nu_1}, \cdots, E'_{\nu_k}$.

REMARK. If in our Definitions 2 and 3 we replace the sign ">" by the sign "≧", then we obtain the inequalities in Theorems 1 and 2 with the sign ">" replaced by "≧". The corresponding set of events thus newly defined will be said to be strongly mutually favorable or indifferent (in the first or second sense).

After this modification, we can include events with the probability 1 in our considerations. Also, the events need no longer be distinct and there may now exist implication relations between events or their conjunctions. This modification is useful for the following application.

4. Consider the divisibility of a random positive integer by the set of positive integers. To each positive integer there corresponds an event, namely the event that the random positive integer is divisible by it. The enumerable set of events

$$E_1, E_2, E_3, E_4, \cdots, E_n, \cdots$$

where E_n = the event of divisibility by n, with the probabilities

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots, \frac{1}{n}, \cdots$$

evidently forms a set of strongly mutually favorable or indifferent events in the second sense.

Again, the enumerable set of events

$$E_2', E_3', E_4', \cdots, E_n', \cdots$$

where E_n' = the event of non-divisibility by n, with the probabilities

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \cdots, \frac{n-1}{n}, \cdots$$

evidently also forms a set of strongly mutually favorable or indifferent events in the second sense.

Hence our Theorem 2 can be applied to both sets and in this way we obtain results which belong properly to the elementary theory of numbers.

We shall content ourselves with indicating a few examples.

Let $\{a_1, \cdots, a_n\}$ denote the least common multiple of the natural numbers $a_1, \cdots, a_n$. Then Theorem 1, when applied to the two sets above, gives respectively

THEOREM 1.1: *Let $a_1, \cdots, a_n$ be any positive integers, then we have, for $k = 1, \cdots, n - 1$*

$$\left(\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_{k+1}}\}}\right)^{\binom{n-1}{k}^{-1}} \geqq \left(\prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_k}\}}\right)^{\binom{n-1}{k-1}^{-1}}.$$

THEOREM 1.2: *Also we have,*

$$\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} \left(1 - \sum_{\nu_1 \leqq \mu_1 \leqq \nu_{k+1}} \frac{1}{a_{\mu_1}} + \sum_{\nu_1 \leqq \mu_1 < \mu_2 \leqq \nu_{k+1}} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^{k+1} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_{k+1}}\}}\right)^{\binom{n-1}{k}^{-1}}$$

$$\geqq \prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} \left(1 - \sum_{\nu_1 \leqq \mu_1 \leqq \nu_k} \frac{1}{a_{\mu_1}} + \sum_{\nu_1 \leqq \mu_1 < \mu_2 \leqq \nu_k} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^k \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_k}\}}\right)^{\binom{n-1}{k-1}^{-1}}$$

A trivial corollary of Theorem 1 is

$$p(12 \cdots n) \geqq p_1 p_2 \cdots p_n .$$

Correspondingly we have

$$1 - \sum_{1 \leqq \mu_1 \leqq n} \frac{1}{a_{\mu_1}} + \sum_{1 \leqq \mu_1 < \mu_2 \leqq n} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^n \frac{1}{\{a_1, \cdots, a_n\}}$$
$$\geqq \left(1 - \frac{1}{a_1}\right)\left(1 - \frac{1}{a_2}\right) \cdots \left(1 - \frac{1}{a_n}\right).$$

If we multiply by $a_1 a_2 \cdots a_n$, we get

$$A(a_1, a_2, \cdots, a_n) \geqq (a_1 - 1)(a_2 - 1) \cdots (a_n - 1),$$

where $A(a_1, \cdots, a_n)$ denotes the number of positive integers $\leqq a_1 a_2 \cdots a_n$ that are not divisible by any of the a_i $(i = 1, \cdots, n)$.

This last result, which is almost obvious here, was proved by H. Rohrbach and H. Heilbronn independently.[2] See also my generalization[3] (also obvious from the present point of view) of this result to higher dimensional sets of positive integers and to sets of ideals in any algebraic number field.

[2] "Beweis einer zahlentheoretische Ungleichung," *Jour. für Math.*, Vol 177 (1937), pp. 193–196 "On an inequality in the elementary theory of numbers," *Proc. Camb. Phil. Soc*, Vol. 33, (1937), pp. 207–209.

[3] "A generalization of an inequality in the elementary theory of numbers," *Jour. für Math.*, Vol. 183 (1941), p. 103.

OBSERVATIONS ON ANALYSIS OF VARIANCE THEORY

By Hilda Geiringer[1]

Bryn Mawr College

One of the important problems of theoretical statistics is the following. Let $x_1, x_2, \cdots x_N$ be the results of N observations; by means of these results we want to test the hypothesis that $V_i(x)$ is the distribution of the ith chance variable x_i. In that situation we often decide to choose a test function $F(x_1, x_2, \cdots x_N)$ and to determine the distribution of F under the above assumption. By means of this distribution we compute the probability of $\xi_1 \leqq F \leqq \xi_2$ and compare this result with the observed value of F.

Suppose there are m groups, each of n observations on $m \cdot n$ chance variables $x_{\mu\nu}$. We may test hypotheses regarding the mn distributions of the $x_{\mu\nu}$ in the way just mentioned. In analysis of variance theory we often use as test functions certain quadratic forms s_w^2 and s_a^2 ("variance within" and "among classes") and their quotient (multiplied by $m(n-1)/(m-1)$), usually denoted by z. Its distribution has been investigated by R. A. Fisher [2] under the assumption that the chance variables are mutually independent and subject to the same normal law. "The five per cent and one per cent points of this distribution have been tabulated by R. A. Fisher and are used to test, whether these two estimates of the same magnitude are significantly different. One gets thus a test of significance *to test whether our sample is a random sample from a homogeneous normal population or not.*[2] If the probability of a certain z-value is too small we shall reject the hypothesis that the sample is a random sample from a homogeneous normal population" [5].

The use of Fisher's z-test is also recommended if we may reasonably assume that the theoretical distributions are approximately normal. "Unless some rather startling lack of normality is known or suspected analysis of variance may be used with confidence." This last remark can be understood by considering that, as we will see in detail, some of the basic results of our theory are independent of the normality of the populations It is however this assumption of normality which makes possible the complete and elegant solution of the problem of distribution obtained by R. A. Fisher.

If it is *not* possible to determine the exact distribution of a test function under sufficiently general assumptions we may:

a) make simple and particular assumptions concerning the populations

b) investigate an asymptotic solution of the problem, i.e. determine the distributions of the test functions for large samples,[3] or

c) study the mathematical expectations and the variances of the test functions

[1] Research under a grant in aid of the American Philosophical Society.

[2] My italics.

[3] cf. statement (a) page 355.

for small samples under appropriately general assumptions regarding the populations (this should be done independently of concepts of estimation, unbiased estimate etc.).
This last procedure provides us with tests which suffice in actual practice.[4]

It is well known that the expectations of the two forms $s_a^2/(m-1)$ and $s_w^2/m(n-1)$ are the same even if the populations are not normal, but equal each other (*Bernoulli* series). In addition we shall prove the theorem, familiar in case of the Lexis quotient [9], that under these conditions the *expectation of their quotients equals unity* (section 1, (b)). The next step consists in investigating certain inequalities characteristic of *Lexis* or *Poisson* series. The different criteria will be completed by the computation of the respective variances (Section 1, (c)).

In addition to the above mentioned test functions other symmetrical test functions have been considered [5] In studying these we shall again assume general populations. It will be seen that the Lexis as well as the Poisson series may be characterized by equalities (instead of inequalities) (Section 2, (a)), and we can generalize our theorem on the expectation of the quotient (Section 2, (b)) to this case. Then the variances of these test functions will be investigated.

It seems worthwhile *to omit the assumption of independence* of the chance variables and to study different kinds of mutual dependence. These investigations lead to interesting relations among the expectations[5] (Section 2, (c)). They seem to be related to Fisher's "intraclass correlation" and to supplement his idea.

Most of the results of Sections 1 and 2 can be generalized to the analysis of covariance (section 3).

1. Variance within and among classes.

(a). The test functions. Let $x_{\mu\nu}$ $(\mu = 1, \cdots m;\ \nu = 1, \cdots n)$ be $m \cdot n$ chance variables and put

$$\text{(1)} \quad \begin{aligned} a_\mu &= \frac{1}{n}\sum_{\nu=1}^{n} x_{\mu\nu}, \qquad \bar{a}_\nu = \frac{1}{m}\sum_{\mu=1}^{m} x_{\mu\nu}, \\ a &= \frac{1}{mn}\sum_{\mu=1}^{m}\sum_{\nu=1}^{n} x_{\mu\nu} = \frac{1}{m}\sum_{\mu=1}^{m} a_\mu = \frac{1}{n}\sum_{\nu=1}^{n} \bar{a}_\nu. \end{aligned}$$

[4] The important paper of Irwin [5] assumes normality of the populations. H L. Rietz [8] computes the expectations of s_a^2 and s_w^2 under rather general assumptions for the populations and considers the cases of Bernoulli, Lexis, Poisson series, but does not consider tests of significance; nor does he consider the symmetric test functions (section 2 of this paper) In later papers on our subject the assumption of normal and independent populations recurs. Another approach [11] in the problem of analysis of variance is to use ranks instead of the actual values (this has been pointed out by the referee to the author, who is very grateful for this comment).

[5] They generalize previous results of the author.

We then introduce the three quadratic forms

$$(2)\quad s^2 = \sum_\mu \sum_\nu (x_{\mu\nu} - a)^2; \qquad s_a^2 = n \sum_\mu (a_\mu - a)^2; \qquad s_w^2 = \sum_\mu \sum_\nu (x_{\mu\nu} - a_\mu)^2,$$

with the respective ranks (degrees of freedom)

$$(3)\qquad r = mn - 1, \qquad r_a = m - 1, \qquad r_w = m(n - 1).$$

Then we have

$$(4)\qquad s^2 = s_a^2 + s_w^2, \qquad r = r_a + r_w.$$

The $m \cdot n$ theoretical distributions are assumed in this section to be independent of each other. Let $V_{\mu\nu}(x)$ be the probability that $x_{\mu\nu} \leqq x$ and

$$(5)\qquad \alpha_{\mu\nu} = \int x\, dV_{\mu\nu}(x), \qquad \sigma_{\mu\nu}^2 = \int (x - \alpha_{\mu\nu})^2\, dV_{\mu\nu}(x),$$

where the integrals are *Stieltjes* integrals; thus the $V_{\mu\nu}(x)$ may be e.g. general arithmetical or geometrical distributions.[6]

Let us compute the mathematical expectation of the three test functions with respect to the $m \cdot n$-dimensional distribution:

$$V_{11}(x_{11})V_{12}(x_{12}) \cdots V_{mn}(x_{mn}).$$

$$(6)\quad E[F(x_{11}, \cdots x_{mn})] = \int \cdots \int F(x_{11}, \cdots x_{mn})\, dV_{11}(x_{11}) \cdots dV_{mn}(x_{mn}).$$

We have then

$$(7)\qquad E\left[\frac{s^2}{mn - 1}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{mn - 1} \Sigma\Sigma(\alpha_{\mu\nu} - \alpha)^2,$$

$$(8)\qquad E\left[\frac{s_{a1}}{m - 1}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{m - 1} \cdot n\Sigma(\alpha_\mu - \alpha)^2,$$

$$(9)\qquad E\left[\frac{s_w^2}{m(n - 1)}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{m(n - 1)} \Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu)^2.$$

From these equalities we deduce:

1. *If the $m \cdot n$ theoretical mean values $\alpha_{\mu\nu}$ are all equal* (Bernoulli series), *then the expectations in* (6), (7), (8) *are equal; i.e.*

$$(10)\qquad E_B\left(\frac{s^2}{mn - 1}\right) = E_B\left(\frac{s_a^2}{m - 1}\right) = E_B\left(\frac{s_w^2}{m(n - 1)}\right).$$

2. *If the $\alpha_{\mu\nu}$ are equal "by rows" but differ from row to row* (Lexis series), i.e. $\alpha_{\mu\nu} = \alpha_\mu$ *but* $\alpha_\mu \neq \alpha$. *Then*

[6] $V_{\mu\nu}(x)$ is a monotone non-decreasing function. Hence it has at most a denumerable set of ordinary jump discontinuities; at such a point it is continuous to the right but not to the left. Moreover it possesses a finite derivative $v_{\mu\nu}(x)$ almost everywhere.

$$(11)\qquad E_L\left[\frac{s_a^2}{m-1}-\frac{s^2}{mn-1}\right]=\frac{mn(n-1)}{(m-1)(mn-1)}\sum_\mu(\alpha_\mu-\alpha)^2>0,$$

$$(12)\qquad E_L\left[\frac{s_a^2}{m-1}-\frac{s_w^2}{m(n-1)}\right]=\frac{n}{m-1}\sum_\mu(\alpha_\mu-\alpha)^2>0.$$

3. *If the $\alpha_{\mu\nu}$ are equal "by columns" but differ from column to column* (Poisson series), *then* $\alpha_{\mu\nu}=\bar{\alpha}_\nu$; $\alpha_\mu=\alpha$ *and*

$$(13)\qquad E_P\left[\frac{s_a^2}{m-1}-\frac{s^2}{mn-1}\right]=-\frac{m}{mn-1}\sum_\nu(\bar{\alpha}_\nu-\alpha)^2<0,$$

$$(14)\qquad E_P\left[\frac{s_a^2}{m-1}-\frac{s_w^2}{m(n-1)}\right]=-\frac{1}{n-1}\sum_\nu(\bar{\alpha}_\nu-\alpha)^2<0.$$

In the Lexis theory[7] we speak of *normal, supernormal* or *subnormal* dispersion depending on whether the *observed* value of $\frac{s_a^2}{m-1}$ is *equal, greater* or *less* than that of $\frac{s^2}{mn-1}$ and we usually consider the quotient

$$(15)\qquad L=\frac{s_a^2}{m-1}\bigg/\frac{s^2}{mn-1}.$$

In analysis of variance theory we usually compare $s_a^2/(m-1)$ (variance among rows) with $s_w^2/m(n-1)$ (variance within rows) and introduce the quotient

$$(16)\qquad V=\frac{s_a^2}{m-1}\bigg/\frac{s_w^2}{m(n-1)}.$$

It follows from (4). *If* $L \gtreqless 1$ *then* $V \gtreqless 1$ *and conversely.* We may therefore speak of *normal* or *non-normal dispersion* with respect either to L or to V.

The results given by equations (10)–(14) can be expressed as follows: *If the $m\cdot n$ theoretical distributions are all equal the mathematical expectation of s^2/r, of s_a^2/r_a and of s_w^2/r_w are equal. In the case of a Lexis series the expectation of s_a^2/r_a is greater than s^2/r and greater than s_w^2/r_w and in the case of a Poisson series the opposite is true.*

We generally use these facts in order to *make inferences about the unknown populations from the observed values of our test functions $V_{\mu\nu}(x)$.* If e.g., the observed value of s_a^2/r_a is "significantly"[8] greater than that of s^2/r we may assume that the theoretical distributions form a Lexis series. But of course such a significant deviation can also be explained by quite different assumptions regarding the populations (see Section 2, (c)).

(*b*). *Mathematical expectation of the quotient of the test functions.* We are going to prove in this section a theorem of some mathematical interest. This theorem is a generalization of an analogous theorem in the Lexis theory [9].

[7] The relation between these considerations and the Lexis theory will be dealt with in another paper.

[8] The meaning of the word "significantly" has of course still to be explained.

We have seen (10) that the mathematical expectations, defined by (6), of the three test functions

$$S = \frac{s^2}{mn - 1}, \qquad S' = \frac{s_a^2}{m - 1}, \qquad S'' = \frac{s_w^2}{m(n - 1)},$$

are equal *if the $m \cdot n$ populations are equal (i.e. have identical distributions). We will show that even in this case*

$$E\left(\frac{S'}{S}\right) = 1, \qquad E\left(\frac{S''}{S}\right) = 1. \tag{17}$$

Let us put $m \cdot n = N$, and let the N chance variables be arranged in a one-dimensional sequence. As S' and S are of second degree in the x_ν ($\nu = 1, 2, \cdots N$) we may write

$$S' - S = A + \sum B_\nu x_\nu + \sum C_\nu x_\nu^2 + \sum_{\nu \neq \rho} D_{\nu\rho} x_\nu x_\rho$$

where the A, B_ν, C_ν and $D_{\nu\rho}$ are constants. Now form the expectation, defined by (6), of $(S' - S)$ under the assumption that the N populations are equal $V_\nu(x) = V(x)$ ($\nu = 1 \cdots N$) Denoting by α and σ^2 the mean value and variance of $V(x)$ and putting $\Sigma B_\nu = B$, $\Sigma C_\nu = C$, $\Sigma D_{\nu\rho} = D$ we find

$$E(S' - S) = A + B\alpha + C(\sigma^2 + \alpha^2) + D\alpha^2 = 0.$$

And as this equality holds for an arbitrary distribution $V(x)$, we deduce that $A = B = C = D = 0$ Let us then compute under the same assumption the expectation of $(S' - S)/S$. Now the expectations of $1/S$, x_ν/S, x_ν^2/S, $x_\nu x_\rho/S$, take the place of the expectations of 1, x_ν, x_ν^2, $x_\nu x_\rho$. But these new expectations are also *independent of the index ν*, because of the *equality of the N populations* and the *symmetry* of S in the N variables $x_1, \cdots x_N$. Hence we may write

$$E\left(\frac{1}{S}\right) = \mu_0, \qquad E\left(\frac{x_\nu}{S}\right) = \mu_1, \qquad E\left(\frac{x_\nu^2}{S}\right) = \mu_2, \qquad E\left(\frac{x_\nu x_\rho}{S}\right) = \mu_3,$$

and we find

$$E\left(\frac{S' - S}{S}\right) = E\left(\frac{S'}{S} - 1\right) = A\mu_0 + B\mu_1 + C\mu_2 + D\mu_3 = 0,$$

because $A = B = C = D = 0$. Hence $E(S'/S) = 1$.

We may prove in the same way that $E(S''/S) = 1$.

We have however proved (17) only under the assumption that all the N *populations are equal*, whereas (10) is true under the mere hypothesis that the mean values of the populations $V_\nu(x)$ are the same.

(c). *The variances of the test functions.* The distribution of our test functions and of their quotients V or L have been determined and tabulated by R. A. Fisher under the hypothesis that the $m \cdot n$ chance variables are *independent* and obey the same normal Gaussian law. Consequently by means of Fisher's distri-

bution we can test only the hypothesis that the theoretical populations have *both these* properties

If in a statistical problem it is not possible to determine the exact distributions of the test functions under sufficiently general assumptions regarding the populations, one of the following procedures is frequently used:
a) one tries to find an *asymptotic* solution of the problem, i.e to determine the distribution of the test functions in question for *large samples*. The distribution of the analysis of variance quotient, as n tends to infinity, has been established by W. G. Madow [6]. The same problem for the Lexis quotient was solved as early as 1873 by Helmert [4]. *As m tends to infinity* the limiting distribution is a Gaussian distribution, which follows from general theorems of v Mises [7].
b) For *small* samples, i.e. if m and n are finite we may determine the expectations and the variances of the test functions for appropriately general populations and establish in this way a test of significance

In this section we shall compute the variances of our test functions. Let us first assume arbitrary but equal populations $V_\nu(x) = V(x)$ and denote by M_i the ith moment about the mean of $V(x)$.

$$\text{(18)} \qquad \begin{aligned} M_i &= \int (x - \alpha)^i \, dV(x), \qquad\qquad (i = 1, 2, \cdots), \\ \alpha &= \int x \, dV(x), \qquad M_2 = \sigma^2 \end{aligned}$$

Then we find immediately the variance of $S = \dfrac{s^2}{mn - 1}$ using a well-known formula for the variance of a sample variance

$$\text{(19)} \qquad \mathrm{Var}\left\{\frac{s^2}{mn-1}\right\} = \mathrm{Var}\left\{\frac{\Sigma\Sigma(x_{\mu\nu} - a)^2}{mn-1}\right\} = \frac{1}{mn}\left\{M_4 - \frac{mn-3}{mn-1} M_2^2\right\}.$$

If we need the analogous variance in case of different populations we let

$$t^2 = \sum_{\rho=1}^{r} (y_\rho - b)^2 \qquad \text{where } b = \frac{1}{r}(y_1 + \cdots + y_r)$$

and let $V_\rho(y)$, $(\rho = 1, \cdots r)$, be r populations, and

$$\beta_\rho = \int y \, dV_\rho(y), \qquad \frac{1}{r}\sum_{\rho=1}^{r} \beta_\rho = \beta,$$

$$\int (y - \beta_\rho)^i \, dV_\rho(y) = \mu_i^{(\rho)}, \qquad (i = 1, 2, \cdots, \rho = 1, 2, \cdots r), \mu_2^{(\rho)} = \sigma_\rho^2.$$

Then the following formula may be used:

$$\text{(20)} \qquad \begin{aligned} \mathrm{Var}\,(t^2) &= \left(\frac{r-1}{r}\right)^2 \sum_{\rho=1}^{r} [\mu_4^{(\rho)} - \sigma_\rho^4] \\ &\quad + 4\frac{r-1}{r}\sum_{\rho=1}^{r} \mu_3^{(\rho)}(\beta_\rho - \beta)^2 + 4\sum_{\rho=1}^{r} \sigma_\rho^2(\beta_\rho - \beta)^2 + \frac{4}{r^2}\sum_{\rho<\tau} \sigma_\rho^2 \sigma_\tau^2. \end{aligned}$$

We may check (20) by putting the $V_\rho(y)$ all equal to $V(y)$ and find

$$\text{Var}\,(t^2) = \frac{r-1}{r}\,[(r-1)\mu_4 - (r-3)\sigma^4], \tag{20'}$$

in accordance with (19).

In order to determine the variance of s_a^2 by means of these formulae we consider $\frac{1}{m}\sum_\mu (a_\mu - a)^2$ as a sample variance. The n distributions in the μth row are $V_{\mu 1}(x), V_{\mu 2}(x), \cdots V_{\mu n}(x)$. Or, if we assume that they are all equal, simply $V(x) = V(x_{\mu\nu})$. Let us put $\frac{1}{n}\,x_{\mu\nu} = z_{\mu\nu}$ and $V(x_{\mu\nu}) = V'(z_{\mu\nu})$, and denote by $W(a_\mu)$ the distribution of the average of the elements in the μth row:

$$W(a_\mu) = \int \cdots \int dV'(z_{\mu 1})\,dV'(z_{\mu 2}) \cdots dV'(z_{\mu, n-1})V'(a_\mu - z_{\mu 1} - \cdots - z_{\mu, n-1}).$$

There is such a distribution for each row, and we have to find the variance of $\sum_\mu (a_\mu - a)^2$ with respect to the combination of these m distributions. In order to be able to apply (20') we need the second and fourth moments of these distributions. We have for the mean value α' of $W(a_\mu)$:

$$\alpha' = n\cdot(\text{mean value of } V') = n\cdot\frac{1}{n}\,\alpha_\mu = \alpha$$

and for the variance μ_2' of $W(a_\mu)$: $\mu_2' = \frac{\sigma^2}{n}$. We still need μ_4'. By repeated use of the formula

$$\begin{aligned}\iint [(x_1 - a_1) + (x_2 - a_2)]^4\,dV(x_1)\,dV(x_2) &\\ = \int (x_1 - a_1)^4\,dV(x_1) &+ \int (x_2 - a_2)^4\,dV(x_2)\\ &+ 6\int (x_1 - a_1)^2\,dV(x_1)\int (x_2 - a_2)^2\,dV(x_2),\end{aligned}$$

and of the fact that $W(a_\mu)$ is simply the distribution of the sum of n variables $z_{\mu\nu}$ we get:

$$\mu_4' = \frac{1}{n^4}\left(n\mathrm{M}_4 + 6\,\frac{n(n-1)}{2}\,\mathrm{M}_2^2\right) = \frac{1}{n^3}\,(\mathrm{M}_4 + 3(n-1)\mathrm{M}_2^2)$$

where M_4 and M_2 are the values introduced in (18).

We now apply (20') and get

$$\text{Var}\,[\Sigma(a_\mu - a)^2] = \frac{m-1}{m}\,[(m-1)\mu_4' - (m-3)\mu_2'].$$

and substituting the values of μ_2' and μ_4', we find by an easy computation the final result:

$$(21)\qquad \mathrm{Var}\left\{\frac{n}{m-1}\,\Sigma(a_\mu - a)^2\right\} = \frac{1}{mn}\,(\mathrm{M}_4 - 3\mathrm{M}_2^2) + \frac{2}{m-1}\,\mathrm{M}_2^2.$$

If we compare this last formula with (19) we see that the right side in (21) is of order $1/m$, whereas that in (19) is of order $1/mn$. Therefore, for sufficiently large values of n, s^2/r will be "more exact" than s_a^2/r_a. In some presentations of the Lexis theory it is implied that the value s_a^2/r_a is to be compared with the theoretical or exact value s^2/r; we may see a certain justification for this idea in the result just mentioned. This may lead us also to use s^2/r as an unbiased estimate of the unknown population variance if $\alpha_{\mu\nu} = \alpha$ (see (7) and (8)).

By means of the simple formulae (19) and (21) we can now easily test whether the values of s^2/r and s_a^2/r_a whose expectations are equal in case of equal populations differ significantly from each other. Of course we must compute as usual approximate values of M_2 and M_4 from the observations If n is comparatively large—as it usually is e.g. in the Lexis theory—only the term $\frac{2}{m-1}\,\mathrm{M}_2^2$ will be significant. If the hypothetical population is Gaussian ($\mathrm{M}_4 = 3\mathrm{M}_2^2$) the right side of (21) reduces to $\frac{2}{m-1}\,\mathrm{M}_2^2$ and that of (19) to $\frac{2\mathrm{M}_2^2}{mn-1}$; hence these variances are in the ratio of $\frac{1}{r_a}\Big/\frac{1}{r}$, as one might expect.

2. Symmetric Test Functions.

(*a*). *New equalities for Lexis and Poisson series.* In Section 1, starting with the formula $s^2 = s_a^2 + s_w^2$ we used the test functions s^2/r, s_a^2/r_a, s_w^2/r_w. This implied a difference between rows and columns, which is often justified, e.g. in the Lexis theory. The following decomposition of s^2 is symmetric with respect to rows and columns. Let

$$(1)\qquad \begin{gathered} \frac{1}{n}\sum_{\nu=1}^{n} x_{\mu\nu} = a_\mu, \qquad \frac{1}{m}\sum_{\mu=1}^{m} x_{\mu\nu} = \bar{a}_\nu, \\ \frac{1}{mn}\sum_{\mu=1}^{m}\sum_{\nu=1}^{n} x_{\mu\nu} = \frac{1}{m}\sum_{\mu=1}^{m} a_\mu = \frac{1}{n}\sum_{\nu=1}^{n}\bar{a}_\nu = a, \end{gathered}$$

and

$$(2)\qquad \begin{aligned} s^2 &= \Sigma\Sigma(x_{\mu\nu} - a)^2, & s_a^2 &= n\Sigma(a_\mu - a)^2, & s_w^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)^2 \\ S^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2, & \bar{s}_a^2 &= m\Sigma(\bar{a}_\nu - a)^2, & \bar{s}_w^2 &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)^2 \end{aligned}$$

with the respective ranks

$$(3)\qquad \begin{aligned} r &= mn - 1, & r_a &= m - 1, & r_w &= m(n-1), \\ R &= (m-1)(n-1), & \bar{r}_a &= n - 1, & \bar{r}_w &= n(m-1). \end{aligned}$$

Then

$$(5)\qquad s^2 = s_a^2 + \bar{s}_a^2 + S^2 = s_a^2 + s_w^2 = \bar{s}_a^2 + \bar{s}_w^2$$

and

$$r = r_a + \bar{r}_a + R = r_a + r_w = \bar{r}_a + \bar{r}_w . \tag{6}$$

We find the expectations of these forms under the assumptions, of arbitrary populations $V_{\mu\nu}(x)$ which are independent and different from each other. We then specialize for Bernouilli series, Lexis and Poisson series of populations respectively. Denoting by $\alpha_{\mu\nu}$ and $\sigma^2_{\mu\nu}$ the mean value and variance of $V_{\mu\nu}(x)$ and by

$$\alpha_\mu = \frac{1}{n}\sum_\nu \alpha_{\mu\nu}, \qquad \bar{\alpha}_\nu = \frac{1}{m}\sum_\mu \alpha_{\mu\nu}, \qquad \alpha = \frac{1}{m}\sum \alpha_\mu = \frac{1}{n}\sum \bar{\alpha}_\nu, \tag{6}$$

we find for the expected values defined in (6) Section 1:

$$\begin{aligned}
E\left[\frac{s^2}{mn-1}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{mn-1}\Sigma\Sigma(\alpha_{\mu\nu}-\alpha)^2,\\
E\left[\frac{s_a^2}{m-1}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{m-1}n\Sigma(\alpha_\mu-\alpha)^2,\\
E\left[\frac{\bar{s}_a^2}{n-1}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{n-1}m\Sigma(\bar{\alpha}_\nu-\alpha)^2,\\
E\left[\frac{S^2}{(m-1)(n-1)}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{(m-1)(n-1)}\Sigma\Sigma(\alpha_{\mu\nu}-\alpha_\mu-\bar{\alpha}_\nu+\alpha)^2,\\
E\left[\frac{s_w^2}{m(n-1)}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{m(n-1)}\Sigma\Sigma(\alpha_{\mu\nu}-\alpha_\mu)^2,\\
E\left[\frac{\bar{s}_w^2}{n(m-1)}\right] &= \frac{1}{mn}\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{n(m-1)}\Sigma\Sigma(\alpha_{\mu\nu}-\bar{\alpha}_\nu)^2.
\end{aligned} \tag{7}$$

In the Bernouilli case which as far as the author knows is the only one which has been considered in this connection [5], we get the wellknown result:

$$\begin{aligned}
&E_B\left[\frac{s^2}{mn-1}\right] = E_B\left[\frac{s_a^2}{m-1}\right] = E_B\left[\frac{\bar{s}_a^2}{n-1}\right]\\
&\quad = E_B\left[\frac{s_w^2}{m(n-1)}\right] = E_B\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = E_B\left[\frac{S^2}{(m-1)(n-1)}\right].
\end{aligned} \tag{8}$$

Now let us assume a Lexis series, with

$$\alpha_{\mu\nu} = \alpha_\mu; \qquad \alpha_\mu \neq \alpha; \qquad \bar{\alpha}_\nu = \alpha, \qquad \sigma^2_{\mu\nu} = \sigma^2_\mu \tag{9}$$

Then (7) reduces to

$$\begin{aligned}
E_L\left[\frac{s^2}{mn-1}\right] &= \frac{1}{m}\Sigma\sigma^2_\mu + \frac{1}{mn-1}n\Sigma(\alpha_\mu-\alpha)^2,\\
E_L\left[\frac{s_a^2}{m-1}\right] &= \frac{1}{m}\Sigma\sigma^2_\mu + \frac{1}{m-1}\cdot n\Sigma(\alpha_\mu-\alpha)^2,\\
E_L\left[\frac{\bar{s}_a^2}{n-1}\right] &= \frac{1}{m}\Sigma\sigma^2_\mu,
\end{aligned}$$

$$E_L\left[\frac{S^2}{(m-1)(n-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2,$$

$$E_L\left[\frac{s_w^2}{m(n-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2,$$

$$E_L\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2 + \frac{1}{m-1}\,\Sigma(\alpha_\mu - \alpha)^2.$$

From these formulae we deduce—besides the inequalities (11), (12) of Section 1, and the corresponding formulae where the role of rows and columns is interchanged—the further inequalities.

$$(11) \qquad E_L\left[\frac{s_a^2}{m-1}\right] > E_L\left[\frac{\bar{s}_w^2}{n(m-1)}\right] > E_L\left[\frac{\bar{s}_a^2}{n-1}\right].$$

But there are also characteristic equalities, namely·

$$(12) \qquad E_L\left[\frac{\bar{s}_a^2}{n-1}\right] = E_L\left[\frac{S^2}{(m-1)(n-1)}\right] = E_L\left[\frac{s_w^2}{m(n-1)}\right].$$

These *equalities*[9] seem often to be more appropriate than the usual *inequalities* in testing the hypothesis of a Lexis series

Let us finally consider the Poisson case which is very often neglected. There we have.

$$(13) \qquad \alpha_{\mu\nu} = \bar{\alpha}_\nu, \qquad \bar{\alpha}_\nu \neq \alpha, \qquad \alpha_\mu = \alpha, \qquad \sigma_{\mu\nu}^2 = \bar{\sigma}_\nu^2.$$

Then—beside the inequalities (13), (14) of Section 1 and the corresponding ones where the role of rows and columns is interchanged—we find the new *inequality*:

$$(14) \qquad E_P\left[\frac{s_a^2}{m-1} - \frac{\bar{s}_a^2}{n-1}\right] = -\frac{m}{n-1}\sum_\nu (\bar{\alpha}_\nu - \alpha)^2 < 0,$$

which of course corresponds to the Lexis inequality (11). The characteristic *equalities* are now:

$$(15) \qquad E_P\left[\frac{s_a^2}{m-1}\right] = E_P\left[\frac{S^2}{(m-1)(n-1)}\right] = E_P\left[\frac{\bar{s}_w^2}{n(m-1)}\right].$$

These equalities (12) and (15) can be used in testing the hypothesis of Lexis or Poisson series respectively in the same way as the equalities (9) for the Bernouilli case. We shall deal with the variances of these test functions in (d) of this section.

(*b*). *Mathematical expectations of the quotients of certain test functions.* We have seen that in case of a Lexis-Series the expectations of $\frac{\bar{s}_a^2}{n-1}$, of $\frac{S^2}{(m-1)(n-1)}$ and of $\frac{s_w^2}{m(n-1)}$ are equal. We will show that even in this case

[9] See [10] pp. 81–90 for proofs of these inequalities for the case of normal populations.

$$
(16)\qquad
\begin{aligned}
E_L\left[\frac{\bar{s}_a^2}{n-1}\Big/\frac{s_w^2}{m(n-1)}\right] &= 1,\\
E_L\left[\frac{s_w^2}{m(n-1)}\Big/\frac{S^2}{(m-1)(n-1)}\right] &= 1,\\
E_L\left[\frac{\bar{s}_a^2}{n-1}\Big/\frac{S^2}{(m-1)(n-1)}\right] &= 1,\\
E_L\left[\frac{S^2}{(m-1)(n-1)}\Big/\frac{s_w^2}{m(n-1)}\right] &= 1.
\end{aligned}
$$

Let us write for the moment: $\frac{\bar{s}_a^2}{n-1} = \bar{T}$ and $\frac{S^2}{(m-1)(n-1)} = T$. As both T and $\bar{T}$ are of second degree in the $x_{\mu\nu}$ we may write:

$$\bar{T} - T = A + \sum_{\mu,\nu} B_{\mu\nu} x_{\mu\nu} + \sum_{\mu,\nu} C_{\mu\nu} x_{\mu\nu}^2 + \sum \sum D_{\mu_1,i;\mu_2,j} x_{\mu_1,i} x_{\mu_2 j},$$

where the A, B, C, D are constants. The last sum contains $\frac{1}{2}\cdot mn(mn - 1)$ terms and *not both* $\mu_1 = \mu_2$ and $i = j$ hold. Compute the expectation of $\bar{T} - T$ with respect to populations which form a Lexis series $V_{\mu\nu}(x) = V_\mu(x)$. Denote by α_μ, σ_μ^2 the respective mean values and variances. We then have because of (11):

$$0 = E_L[\bar{T} - T] = A + \sum_\mu \alpha_\mu \sum_\nu B_{\mu\nu} + \sum_\mu (\sigma_\mu^2 + \alpha_\mu^2) \sum_\nu C_{\mu\nu} + \sum_{\mu_1,\mu_2} \alpha_{\mu_1} \alpha_{\mu_2} \sum_{i,j} D_{\mu_1 i;\mu_2 j}$$

or introducing $\sum_\nu B_{\mu\nu} = B_\mu$; $\sum_\nu C_{\mu\nu} = C_\mu$; $\sum_{i,j} D_{\mu_1 i;\mu_2 j} = D_{\mu_1\mu_2}$ we get:

$$0 = E_L[\bar{T} - T] = A + \sum_\mu \alpha_\mu B_\mu + \sum_\mu (\sigma_\mu^2 + \alpha_\mu^2) C_\mu + \sum_{\mu_1,\mu_2} \alpha_{\mu_1} \alpha_{\mu_2} D_{\mu_1\mu_2}.$$

As this equality is exact for an arbitrary set of $V_\mu(x)$ we deduce that $A = 0$, $B_\mu = 0$, $C_\mu = 0$, $D_{\mu_1\mu_2} = 0$.

Let us now compute under the same assumption the expectation of $(\bar{T} - T)/T$. Here the expectations of $1/T$, $x_{\mu\nu}/T$ etc. will take the place of the expectations of 1, $x_{\mu\nu}$, $\cdots$. But these new expectations will not depend on the index ν (index within the row) because the populations are the same within each row and because of the symmetry of T in the $m \cdot n$ variables $x_{\mu\nu}$. Hence we can put

$$E\left(\frac{1}{T}\right) = l_0, \qquad E\left(\frac{x_{\mu\nu}}{T}\right) = l_\mu, \qquad E\left(\frac{x_{\mu\nu}^2}{T}\right) = l_\mu, \qquad E\left(\frac{x_{\mu_1 i} x_{\mu_2 j}}{T}\right) = l_{\mu_1\mu_2}, \quad \text{etc.}$$

and we get

$$E\left[\frac{\bar{T} - T}{T}\right] = E\left(\frac{\bar{T}}{T} - 1\right) = Al_0 + \sum_\mu l_\mu B_\mu + \sum_\mu l_\mu C_\mu + \sum_{\mu_1,\mu_2} l_{\mu_1,\mu_2} D_{\mu_1,\mu_2} = 0,$$

because all the coefficients are equal to zero. Our theorem is thus proved. The same conclusion holds if the denominator—without being symmetric in all the

$m \cdot n$ variables—does not depend on the row index. And as this last property holds for s_w^2 the expectations (16) are all shown to be equal to one.

Analogous relations are valid for Poisson series.

(c). *Non-independent populations.* We omit in this section the assumption of independence of the $m \cdot n$ populations but assume the theoretical population to be a general $m \cdot n$-variate distribution:

$$V(x_{11}, x_{12}, \cdots x_{mn}). \tag{17'}$$

From $V(x_{11}, x_{12}, \cdots x_{mn})$ we derive the mn one-dimensional distributions $V_{\mu\nu}(x)$ ($\mu = 1, \cdots m$; $\nu = 1, \cdots n$) by letting all the variables except $x_{\mu\nu}$ tend to $+\infty$, because $V_{\mu\nu}(x)$ is the probability that $x_{\mu\nu} \leqq x$ regardless of the values of the other variables. In a similar way we derive the $\frac{1}{2}mn(mn - 1)$ two dimensional distributions $V_{\mu_1\nu_1;\mu_2\nu_2}(x, y)$, that is the probability that $x_{\mu_1\nu_1} \leqq x$ and $x_{\mu_2\mu_2} \leqq y$. We get this distribution from (17′) as all the variables with the exception of $x_{\mu_1\nu_1}$ and $x_{\mu_2\nu_2}$ tend to $+\infty$. We denote as before by $\alpha_{\mu\nu}$ and $\sigma_{\mu\nu}^2$ the expectation of $x_{\mu\nu}$ and $(x_{\mu\nu} - \alpha_{\mu\nu})^2$ respectively. But the expectation of $(x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})$ which was zero in case of the independence of $x_{\mu_1\nu_1}$ and $x_{\mu_2\nu_2}$ may now differ from zero Denote by $\mathfrak{E}$ the expectation with respect to (17′). Then:

$$\begin{aligned} &\mathfrak{E}[(x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})] \\ &\quad = \int\int \cdots \int (x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})\, dV(x_{11}, \cdots x_{mn}) \\ &\quad = \int\int (x - \alpha_{\mu_1\nu_1})(y - \alpha_{\mu_2\nu_2})\, dV_{\mu_1\nu_1,\nu_2\mu_2}(xy) = R_{\mu_1\nu_1;\mu_2\nu_2} = R_{\mu_2\nu_2,\mu_1\nu_1}. \end{aligned} \tag{17}$$

Let us first deduce a general formula for the expectation of a sample variance in the case of dependent populations. Let $P(y_1, \cdots y_r)$ be the distribution of r chance variables $y_1, \cdots y_r$ which have the average b. Denoting by β_ρ the expectation of y_ρ with respect to P, by β the average of the β_ρ, by τ_ρ^2 the expectation of $(y_\rho - \beta_\rho)^2$ by R_{ij}, that of $(y_i - \beta_i)(y_j - \beta_j)$ we find, without difficulty, for the expectation of the sample variance

$$\begin{aligned} &\text{Exp.}\left[\frac{1}{r}\sum_{\rho=1}^{r}(y_\rho - b)^2\right] \\ &\quad = \frac{1}{r}\int \cdots \int [(y_1 - b)^2 + \cdots + (y_r - b)^2]\, dP(y_1, \cdots y_r) \\ &\quad = \frac{r-1}{r^2}\sum_{\rho=1}^{r}\tau_\rho^2 + \frac{1}{r}\sum_{\rho}(\beta_\rho - \beta)^2 - \frac{2}{r^2}\sum_{i<j} R_{ij}. \end{aligned} \tag{18}$$

Let us apply this result in the computation of the expectations of our test functions. It is not difficult to compute them in the general case of *different* mean values and variances But we restrict ourselves to the consideration of certain particular cases. Take first the case where all the $m \cdot n$ mean values $\alpha_{\mu\nu}$ are equal

to each other and likewise the $m \cdot n$ variances and the $\frac{1}{2}mn(mn - 1)$ covariances. Denote these magnitudes by α, σ^2 and R, respectively, we see from (18) that:

$$\mathfrak{E}\left(\frac{s^2}{mn-1}\right) = \mathfrak{E}\left(\frac{s_a^2}{m-1}\right) = \mathfrak{E}\left(\frac{\bar{s}_a^2}{n-1}\right)$$

(19)

$$= \mathfrak{E}\left(\frac{s_w^2}{m(n-1)}\right) = \mathfrak{E}\left(\frac{\bar{s}_w^2}{n(m-1)}\right) = \mathfrak{E}\left(\frac{S^2}{(m-1)(n-1)}\right)$$

$$= \sigma^2 - R.$$

We have thus obtained the result *that in the case of dependent populations, just described, the expectations of the six different test functions are still the same.*

Of course we may assume many other particular kinds of mutual dependence of the populations. The following assumption seems to be appropriate for problems where rows and columns play a *different* role: We consider dependence *only within each* row, that means we assume only the variables $x_{\mu 1}, x_{\mu 2}, \cdots x_{\mu n}$ as mutually dependent. The distribution (16) has then the following form:

$$(20) \qquad V(x_{11}, \cdots x_{mn}) = V_1(x_{11}, \cdots x_{1n})V_2(x_{21}, \cdots x_{2n}) \cdots V_m(x_{m1}, \cdots x_{mn}).$$

In the usual way we derive the $m \cdot n$ one dimensional distributions $V_{\mu\nu}(x)$ and the $\frac{1}{2}mn(mn - 1)$ two-dimensional distributions $V_{\mu_1\nu_1,\mu_2\nu_2}(x, y)$. If $\mu_1 \neq \mu_2$ such a two-dimensional distribution reduces to the product of the respective one-dimensional distributions. Only the $\frac{1}{2}mn(n - 1)$ bivariate distributions derived from one and the same $V_\mu(x_{\mu 1}, \cdots x_{\mu n})$ will not reduce in this way.

Denoting again by $\mathfrak{E}$ the expectation with respect to $V(x_{11}, \cdots x_{mn})$ we find:

$$(21) \qquad \begin{aligned} \mathfrak{E}[(x_{\mu_1 i} - \alpha_{\mu_1 i})(x_{\mu_2 j} - \alpha_{\mu_2 j})] &= 0 && \mu_1 \neq \mu_2 \\ &= R_{ij}^{(\mu_1)} && \mu_1 = \mu_2 \text{ and } i \neq j. \end{aligned}$$

Applying now formula (18) in the computation of the expectations of s^2, s_w^2 and s_a^2 we find:

$$\mathfrak{E}[\sum\sum(x_{\mu\nu} - a)^2] = \frac{mn-1}{mn}\sum\sum\sigma_{\mu\nu}^2 + \sum\sum(\alpha_{\mu\nu} - \alpha)^2 - \frac{2}{mn}\sum_{\mu=1}^{m}\sum_{i<j}R_{ij}^{(\mu)},$$

$$(22) \qquad \mathfrak{E}[\sum\sum(x_{\mu\nu} - a_\mu)^2] = \frac{m(n-1)}{mn}\sum\sum\sigma_{\mu\nu}^2 + \sum\sum(\alpha_{\mu\nu} - \alpha_\mu)^2 - \frac{2}{n}\sum_{\mu=1}^{m}\sum_{i<j}R_{ij}^{(\mu)},$$

$$\mathfrak{E}[\sum\sum(a_\mu - a)^2] = \frac{m-1}{mn}\sum\sum\sigma_{\mu\nu}^2 + n\sum(\alpha_\mu - \alpha)^2 + \frac{2(m-1)}{mn}\sum_{\mu=1}^{m}\sum_{i<j}R_{ij}^{(\mu)}.$$

Let us now suppose that *all the m·n distributions are equal* to each other, or, at least, that·

$$\alpha_{\mu\nu} = \alpha. \tag{23}$$

This assumption, which is characterized by (21), is, of course, different from the one which leads us to (19). We find now by means of (22), if we set

$$\sum_{\mu=1}^{m} \sum_{i<j} R^{\mu}_{ij} = \bar{R} \quad \text{and} \quad \frac{2}{mn(mn-1)} \bar{R} = R,$$

$$\mathfrak{E}\left[\frac{s_a^2}{m-1} - \frac{s^2}{mn-1}\right] = \frac{2}{mn-1} \bar{R} = \frac{mn(n-1)}{mn-1} R. \tag{24}$$

Assuming $R > 0$ (positive average correlation) we may compare this result with (11) Section 1. The term on the right side of (24) is also of the same order of magnitude as that in (11) —For negative R the term on the right side of (24) is negative and the equation may be compared with (13) Section 1. We see that for the test functions s^2/r and s_a^2/r_a *"positive, (negative) average correlation within rows" has the same effect as "Lexis (Poisson) Series" of populations.*

Consider now the test functions $\bar{s}_a^2$ and S^2. We find

$$\mathfrak{E}[\bar{s}a^2] = \mathfrak{E}[\Sigma\Sigma(\bar{a}_\nu - a)^2] = \frac{n-1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + m\Sigma(\bar{\alpha}_\nu - \alpha)^2 - \frac{2}{mn}\bar{R}, \tag{25}$$

and

$$\begin{aligned} \mathfrak{E}[S^2] = \mathfrak{E}[\Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2] &= \frac{(m-1)(n-1)}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 \\ &+ \Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu - \bar{\alpha}_\nu + \alpha)^2 - \frac{2(m-1)}{mn}\bar{R}. \end{aligned} \tag{25'}$$

Assuming (23) we get:

$$\mathfrak{E}\left[\frac{s_a^2}{m-1} - \frac{\bar{s}_a^2}{n-1}\right] = nR, \tag{26}$$

and if $R > 0$:

$$\mathfrak{E}\left[\frac{s_a^2}{m-1}\right] > \mathfrak{E}\left[\frac{\bar{s}_w^2}{n(m-1)}\right] > \mathfrak{E}\left[\frac{\bar{s}_a^2}{n-1}\right]. \tag{26'}$$

The first equality is analogous to (11) and (14) of Section 2 for positive or negative R respectively.[10] We also get under the assumption (23)

$$\mathfrak{E}\left[\frac{\bar{s}_a^2}{n-1}\right] = \mathfrak{E}\left[\frac{S^2}{(m-1)(n-1)}\right] = \mathfrak{E}\left[\frac{s_w^2}{m(n-1)}\right]. \tag{27}$$

[10] I have studied in another paper the *combination* of Lexis series and "positive correlation within rows." It turns out that the two kinds of positive effects reinforce each other. The same is true for "negative correlation" and Poisson series See [3].

These are the same equations as (12) Section 2, and they are true for either sign of R. Hence they provide no way to decide between Lexis series and correlated populations. But computing the expectations of the magnitudes which occur in (15) Section 2 we find from (22), (25) and (25′)

$$\mathfrak{E}\left[\frac{s_a^2}{m-1}\right] = \sigma^2 + (n-1)R, \qquad \mathfrak{E}\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = \sigma^2 \tag{28}$$
$$\mathfrak{E}\left[\frac{S^2}{(m-1)(n-1)}\right] = \sigma^2 - R.$$

And hence we may say:
If the observed value of $s_a^2/(m-1)$ *is greater than that of* $\bar{s}_w^2/n(m-1)$ *this can be explained either by the assumption of a Lexis series or a positive correlation within rows; but their equality indicate, a Poisson series; and if the first is smaller than the second we may assume negative correlation.*
In the same way we may explain

$$\left[\frac{\bar{s}_w^2}{n(m-1)}\right]_{\text{observed}} > \left[\frac{S^2}{(m-1)(n-1)}\right]_{\text{observed}},$$

either by positive correlation or by Lexis series; whereas the equality indicates a Poisson series and the sign $<$ indicates negative correlation.

(*d*). *The variances of the test functions.* We have still to find the variance of our test functions. Let us compute the variance of

$$\Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2$$

with respect to the $m \cdot n$ dimensional distribution $V(x_{11})V(x_{12}) \cdots V(x_{mn})$. Let us put

$$x_{\mu\nu} - a_\mu - \bar{a}_\nu + a = y_{\mu\nu}, \tag{29}$$

then we see that the average of the $y_{\mu\nu}$ equals zero

$$\bar{y} = \frac{1}{mn}\Sigma\Sigma y_{\mu\nu} = a - \frac{1}{mn}n\Sigma a_\mu - \frac{1}{mn}m\Sigma\bar{a}_\nu + a = 0.$$

and

$$S^2 = \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2 = \Sigma\Sigma(y_{\mu\nu} - \bar{y})^2.$$

Each $y_{\mu\nu}$ is a linear function of the $x_{\mu\nu}$ e.g.

$$y_{11} = x_{11}\frac{(m-1)(n-1)}{mn}$$
$$- \frac{m-1}{mn}\sum_{j=2}^{n} x_{1j} - \frac{n-1}{mn}\sum_{i=2}^{m} x_{i1} + \frac{1}{mn}\sum_{2}^{m}\sum_{2}^{n} x_{ij} \tag{30}$$
$$= x_{11}\lambda_1 + \lambda_2\sum_{2}^{n} x_{1j} + \lambda_3\sum_{2}^{m} x_{i1} + \lambda_4\sum_{2}^{m}\sum_{2}^{n} x_{ij}.$$

Using the same notations as in Section 1 (c) we find, because of the independence of each chance variable

$$\text{Var}\,(y_{11}) = \lambda_1^2\sigma^2 + \lambda_2^2(n-1)\sigma^2 + \lambda_3^2(m-1)\sigma^2 + \lambda_4^2(m-1)(n-1)\sigma^2 = \frac{(m-1)(n-1)}{mn}\sigma^2 \tag{31'}$$

and we find the same result for each $y_{\mu\nu}$:

$$\sigma'^2 = \text{Var}\,(y_{\mu\nu}) = \frac{(m-1)(n-1)}{mn}\sigma^2 \tag{31}$$

in agreement with the fourth line of (7) of this section. We still need M_4' the fourth moment about the mean of $y_{\mu\nu}$ which we can compute from the fourth moment of a sum. We find

$$M_4' = AM_4 + 6B\sigma^4, \tag{32}$$

and we have

$$\begin{aligned} A &= \lambda_1^4 + (n-1)\lambda_2^4 + (m-1)\lambda_3^4 + (m-1)(n-1)\lambda_4^4 \\ &= \frac{(m-1)(n-1)}{m^3n^3}(m^2-3m+3)(n^2-3n+3), \end{aligned} \tag{33}$$

and

$$\begin{aligned} B = {} & \lambda_1^2\{\lambda_2^2(n-1) + \lambda_3^2(m-1) + \lambda_4^2(m-1)(n-1)\} \\ & + \lambda_2^2(n-1)\{\tfrac{1}{2}\lambda_2^2(n-2) + \lambda_3^2(m-1) + \lambda_4^2(m-1)(n-1)\} \\ & + \lambda_3^2(m-1)\{\tfrac{1}{2}\lambda_3^2(m-2) + \lambda_4^2(m-1)(n-1)\} \\ & + \tfrac{1}{2}\lambda_4^4(m-1)(n-1)[(m-1)(n-1)-1]. \end{aligned} \tag{34'}$$

If we introduce the values of λ_1, λ_2, λ_3, λ_4 we find

$$\begin{aligned} m^4n^4B = {} & (m-1)^3(n-1)^3(m+n) + (m-1)^2(n-1)^2(m+n-2) \\ & + \tfrac{1}{2}(m-1)(n-1)[(m-1)^3(n-2) + (n-1)^3(m-2) \\ & + (mn-m-n)] \end{aligned} \tag{34}$$

this expression as well as that of A may be easily computed for different values of m and n.

If m and n are large, B is of order $\frac{1}{m} + \frac{1}{n}$; from (31)-(34) we see that in this case σ'^2 is approximately equal to σ^2 and M_4' to M_4.

Using now (18′) we find finally

$$\text{Var}\,\{\Sigma\Sigma\,(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2\} = \frac{mn-1}{mn}\{(mn-1)M_4' - (mn-3)\sigma'^4\}$$

where M_4' and σ'^2 are the expressions just computed. If we compare the variances of the test functions $s_a^2/(m-1)$ and $S^2/(m-1)(n-1)$ we see that whereas the variance of the first expression is of order $1/m$ that of the second is of order $1/mn$. Hence for large values of n the *latter expression is more exact* than the former (see the analogous remark Section 1 (c)). A similar statement can be made if $\bar{s}_a^2/(n-1)$ takes the place of $s_a^2/(m-1)$.

3. Bivariate distributions. Analysis of covariance.

(*a*). *Problem.* Suppose m persons are throwing two dice, n times; we observe the respective numbers on each die in these $m \cdot n$ trials. Or we observe on m groups of n persons the color of the hair and of the eyes. Or else we state for n years the yield of wheat (in bushels) per acre and the production cost (per bushel) for m farms; etc.

We consider $m \cdot n$ pairs of numbers $x_{\mu\nu}$, $y_{\mu\nu}$. Let $V_{\mu\nu}(x, y)$[11] be the probability that $x_{\mu\nu} \leqq x$ and $y_{\mu\nu} \leqq y$; $V_{\mu\nu}(x, +\infty) = V_{\mu\nu}^{(1)}(x)$, $V_{\mu\nu}(+\infty, y) = V_{\mu\nu}^{(2)}(y)$ and introduce the following mean values and variances

$$(1) \qquad \iint x\, dV_{\mu\nu}(x, y) = \alpha_{\mu\nu}, \qquad \iint y\, dV_{\mu\nu}(x, y) = \beta_{\mu\nu},$$

$$(2) \qquad \iint (x-\alpha_{\mu\nu})^2\, dV_{\mu\nu}(x, y) = \sigma_{\mu\nu}^2, \qquad \iint (y-\beta_{\mu\nu})^2\, dV_{\mu\nu}(xy) = \tau_{\mu\nu}^2,$$

$$(3) \qquad \iint (x-\alpha_{\mu\nu})(y-\beta_{\mu\nu})\, dV_{\mu\nu}(x, y) = \gamma_{\mu\nu}$$

$$(4) \qquad \begin{aligned} &\frac{1}{n}\sum_{\nu} \alpha_{\mu\nu} = \alpha_{\mu}, \qquad \frac{1}{m}\sum_{\mu} \alpha_{\mu\nu} = \bar{\alpha}_{\nu}, \qquad \frac{1}{mn}\sum\sum \alpha_{\mu\nu} = \alpha \\ &\frac{1}{n}\sum_{\nu} \beta_{\mu\nu} = \beta_{\mu}, \qquad \frac{1}{m}\sum_{\mu} \beta_{\mu\nu} = \bar{\beta}_{\nu}, \qquad \frac{1}{mn}\sum\sum \beta_{\mu\nu} = \beta \end{aligned}$$

Let us compute the mathematical expectations of certain test functions with respect to the $2mn$-dimensional distributions

$V_{11}(x_{11}, y_{11})V_{12}(x_{12}, y_{12}) \cdots V_{mn}(x_{mn}, y_{mn})$. Let

$$(5) \qquad \begin{aligned} &E[F(x_{11}, y_{11}, \cdots x_{mn}, y_{mn})] \\ &\qquad = \int \cdots \int F(x_{11}, \cdots y_{mn})\, dV_{11}(x_{11}, y_{11}) \cdots dV_{mn}(x_{mn} y_{mn}) \end{aligned}$$

---

[11] In the particular case where $V_{\mu\nu}(x, y)$ has everywhere a derivative $\dfrac{\partial^2 V_{\mu\nu}}{\partial x\, \partial y}$ we can use the two dimensional density $v_{\mu\nu}(x, y) = \dfrac{\partial^2 V_{\mu\nu}}{\partial x\, \partial y}$ and the one-dimensional densities

$$v_{\mu\nu}^{(1)}(x) = \int v_{\mu\nu}(xy)\, dy; \qquad v_{\mu\nu}^{(2)}(y) = \int v_{\mu\nu}(x, y)\, dx$$

and we have

$$V_{\mu\nu}^{(1)}(x) = \int_{-\infty}^{x} v_{\mu\nu}^{(1)}(x)\, dx, \qquad V_{\mu\nu}^{(2)}(y) = \int_{-\infty}^{y} v_{\mu\nu}^{(2)}(y)\, dy.$$

We then have[12]

$$(5') \quad F[G(x_{11}, \cdots x_{mn})] = \int_{(x_{11})} \cdots \int_{(x_{mn})} G(x_{11} \cdots x_{mn})\, dV_{11}^{(1)}(x_{11}) \cdots dV_{mn}^{(1)}(x_{mn}).$$

In analogy with previous notations we introduce

$$(6) \quad \begin{aligned} a_\mu &= \frac{1}{n}\sum_\nu x_{\mu\nu}, & \bar{a}_\nu &= \frac{1}{m}\sum_\mu x_{\mu\nu}, & a &= \frac{1}{mn}\sum\sum x_{\mu\nu}, \\ b_\mu &= \frac{1}{n}\sum_\nu y_{\mu\nu}, & \bar{b}_\nu &= \frac{1}{m}\sum_\mu y_{\mu\nu}, & b &= \frac{1}{mn}\sum\sum y_{\mu\nu}, \end{aligned}$$

and

$$(7) \quad \begin{aligned} s^2 &= \Sigma\Sigma(x_{\mu\nu} - a)^2, & s_a^2 &= n\Sigma(a_\mu - a)^2, & s_w^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)^2 \\ S^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2, & \bar{s}_a^2 &= m\Sigma(\bar{a}_\nu - a)^2, & \bar{s}_w^2 &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)^2 \\ t^2 &= \Sigma\Sigma(y_{\mu\nu} - b)^2, & t_a^2 &= n\Sigma(b_\mu - b)^2, & t_w^2 &= \Sigma\Sigma(y_{\mu\nu} - b_\mu)^2 \\ T^2 &= \Sigma\Sigma(y_{\mu\nu} - b_\mu - \bar{a}_\nu + b)^2, & \bar{t}_a^2 &= m\Sigma(\bar{b}_\nu - b)^2, & \bar{t}_w^2 &= \Sigma\Sigma(y_{\mu\nu} - \bar{b}_\nu)^2, \end{aligned}$$

and

$$(8) \quad \begin{aligned} c &= \Sigma\Sigma(x_{\mu\nu} - a)(y_{\mu\nu} - b), & C &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)(y_{\mu\nu} - b_\mu - \bar{b}_\nu + b) \\ c_a &= n\Sigma(a_\mu - a)(b_\mu - b) & c_w &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)(y_{\mu\nu} - b_\mu) \\ \bar{c}_a &= m\Sigma(\bar{a}_\nu - a)(\bar{b}_\nu - b) & \bar{c}_w &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)(y_{\mu\nu} - \bar{b}_\nu) \end{aligned}$$

we then have

$$(9) \quad \begin{aligned} s^2 &= S^2 + s_a^2 + \bar{s}_a^2 = s_w^2 + s_a^2 = \bar{s}_w^2 + \bar{s}_a^2, \\ t^2 &= T^2 + t_a^2 + \bar{t}_a^2 = t_w^2 + t_a^2 = \bar{t}_w^2 + \bar{t}_a^2, \\ c &= C + c_a + \bar{c}_a = c_a + c_w = \bar{c}_a + \bar{c}_w, \end{aligned}$$

and corresponding relations for the ranks of these quadratic forms. We find for the expectations of these test functions, in analogy with previously investigated formulae:

$$E\left[\frac{t^2}{mn - 1}\right] = \frac{1}{mn}\Sigma\Sigma\tau_{\mu\nu}^2 + \frac{1}{mn - 1}\Sigma\Sigma(\beta_{\mu\nu} - \beta)^2,$$

$$E\left[\frac{t_a^2}{m - 1}\right] = \frac{1}{mn}\Sigma\Sigma\tau_{\mu\nu}^2 + \frac{1}{m - 1} n\Sigma(\beta_\mu - \beta)^2,$$

. . . .

and

$$E\left[\frac{c}{mn - 1}\right] = \frac{1}{mn}\Sigma\Sigma\gamma_{\mu\nu} + \frac{1}{mn - 1}\Sigma\Sigma(\alpha_{\mu\nu} - \alpha)(\beta_{\mu\nu} - \beta),$$

$$E\left[\frac{c_a}{m - 1}\right] = \frac{1}{mn}\Sigma\Sigma\gamma_{\mu\nu} + \frac{1}{m - 1} \cdot n\Sigma(\alpha_\mu - \alpha)(\beta_{\mu\nu} - \beta),$$

. . . .

[12] It may be mentioned that the problem considered in this section of mn bivariate distribution $v_{\mu\nu}(x, y)$ constitutes, of course, only a particular case of dependence (see section 2, (c)) for a $2mn$ dimensional population $v(x_{11}, y_{11}, x_{12}, y_{12}, \cdots x_{mn}, y_{mn})$.

1) *If all the $\alpha_{\mu\nu}$ equal each other, or all the $\beta_{\mu\nu}$ equal each other, we find:*

$$E_B\left[\frac{c}{mn-1}\right] = E_B\left[\frac{c_a}{m-1}\right] = E_B\left[\frac{c_w}{m(n-1)}\right]$$

$$= E_B\left[\frac{C}{(m-1)(n-1)}\right] = E_B\left[\frac{\bar{c}_a}{n-1}\right] = E_B\left[\frac{\bar{c}_w}{n(m-1)}\right] = \frac{1}{mn}\Sigma\Sigma\gamma_{\mu\nu}.$$

These formulae provide us with unbiased estimates of $\Sigma\Sigma\gamma_{\mu\nu}$.

2) *The $\alpha_{\mu\nu}$ are equal within each row but differ from row to row,* (Lexis) $\alpha_{\mu\nu} = \alpha_\mu \neq \alpha$; $\bar{\alpha}_\nu = \alpha$ *whereas the $\beta_{\mu\nu}$ may have arbitrary values, then*

$$E_L\left[\frac{\bar{c}_a}{n-1}\right] = E_L\left[\frac{c_w}{m(n-1)}\right] = E_L\left[\frac{C}{(m-1)(n-1)}\right]. \tag{13}$$

The same equalities are valid for arbitrary $\alpha_{\mu\nu}$ if the $\beta_{\mu\nu} = \beta_\mu$; $\bar{\beta}_\nu = \beta$. Our new equalities may be of some interest because inequalities analogous to those of the Lexis case cannot be proved for covariances. If the observed values of the expressions in (13) are significantly different we may conclude that neither the $\alpha_{\mu\nu}$ nor the $\beta_{\mu\nu}$ form a Lexis series. A judgment of the test (13) might be based on the investigation of its power function. But besides we have the equalities (12) and analogous equalities containing $\bar{t}_a^2$, T^2 and $\bar{t}_w^2$.

3) *If either* $\alpha_{\mu\nu} = \bar{\alpha}_\nu$, $\bar{\alpha}_\nu \neq \alpha$, $\alpha_\mu = \alpha$,

or $\beta_{\mu\nu} = \bar{\beta}_\nu$, $\beta_\nu \neq \beta$, $\beta_\mu = \beta$.

We have the new equalities

$$E_P\left[\frac{c_a}{m-1}\right] = E_P\left[\frac{\bar{c}_w}{n(m-1)}\right] = E_P\left[\frac{C}{(m-1)(n-1)}\right], \tag{14}$$

and there are no inequalities analogous to the inequalities (14) *of Section* 2, *and* (13), (14) *of Section* 1.

Most of the investigations of Sections 1 and 2 can be generalized for this two dimensional problem.

BIBLIOGRAPHY

[1] R. A. Fisher, *Statistical Methods for Research Workers*, 6th ed., p. 214 ff

[2] R. A. Fisher, "Applications of 'Student's' distributions," *Metron*, Vol. 5 (1926), pp. 90–104.

[3] H. Geiringer, "A new explanation of non-normal dispersion in the Lexis theory," *Econometrica*, Vol. 10 (1942), pp. 53–60.

[4] F. R. Helmert, *Zeits. für Math. und Physik*, Vol. 21 (1876), p. 192–218.

[5] I. O. Irwin, "Mathematical theorems involved in the analysis of variance," *Jour. Roy. Stat. Soc.*, Vol 94 (1931), pp. 284–300.

[6] W. G. Madow, "Limiting distributions of quadratic and bilinear forms," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 125–147.

[7] R. v. Mises, "Theorie des probabilites. Fondements et applications," *Annales de l'Institut Poincare*, (1931), pp. 137–190.

[8] H. L. RIETZ, "On the Lexis theory and the analysis of variance," *Bull Am. Math Soc*, (1932), pp 731 ff.

[9] A. A TSCHUPROW, *Skandinavisk Aktuarietidskrift*, Vol. 6 (1918).

[10] A WALD, *Lectures on the Analysis of Variance and Covariance*, *Columbia University*, 1941

[11] MILTON FRIEDMAN, "The use of ranks to avoid the assumption of normality," *Jour. Amer Stat Assn.*, Vol. 32 (1937), pp. 675–701.

THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIII, No. 4 — December, 1942

ON THE RATIO OF THE VARIANCES OF TWO NORMAL POPULATIONS

By Henry Scheffé

Princeton University

CONTENTS

1. Introduction and summary. Suppose that we have two samples E_1 and E_2 from normal populations π_1 and π_2 with unknown means and variances. Let us designate by θ the ratio of the variance of π_1 to that of π_2 *The two problems discussed in this paper are to formulate in terms of E_1 and E_2, and to compare,*

(*i*) *significance tests for the hypothesis that the unknown ratio θ is equal to a given positive number θ_0*, and

(*ii*) *confidence intervals for θ.*

Since, on the one hand, these problems are of considerable importance to the practical statistician and the teacher of statistics, and on the other, they cry for the application of recently developed theory which is unfortunately not yet familiar to many practical workers and teachers, the development has been divided into two parts: Part I, it is hoped, will be intelligible to the above class of readers; part II, slanted toward a smaller circle, is more esoteric, general, and condensed.

More specifically, in part I it is pointed out that any choice of limits on the F-distribution satisfying the condition that the sum of the areas in the tails be equal to a prescribed number, leads to solutions of problems (*i*) and (*ii*). After considering and then ruling out the "one-sided" situations in which it is appropriate to use only one tail, two conditions are proposed (*ad hoc* and on an intuitive basis) for the "two-sided" case,—a symmetry condition, and a condition for logarithmically shortest confidence intervals. The second condition leads to a choice of limits on the F-distribution. From other considerations,—

reciprocal limits, likelihood ratio, and equal tails, other choices are advanced. It is found that all four of these choices satisfy the first condition, and that furthermore if $N_1 = N_2$, where N_i is the number of variates in E_i, then the four choices become identical. If $N_1 \neq N_2$ which of the four tests is "best"? which of the four sets of confidence intervals? For defining and answering the first question in a logically satisfactory way just a little of the Neyman-Pearson theory of testing hypotheses suffices. For the second, Neyman's theory of confidence intervals is called for, and because of its greater difficulty, this has been relegated to part II. However, the limits determined by the criterion that the test be unbiased turn out to be the same as those which yield optimum confidence intervals from the elementary viewpoint of §5. Their numerical values are unfortunately laborious to calculate accurately if $N_1 \neq N_2$, and part I concludes with some numerical evidence indicating the loss of efficiency in using instead the easily found "equal tails" limits. For N_1 and $N_2 \geqq 10$ this loss is seen to be quite small. It will perhaps bear repeating that if $N_1 = N_2$, the "equal tails" limits on the F-distribution are the same as those associated with the unbiased test and that hence in this case all the advantages uncovered in parts I and II for the unbiased test and the related confidence intervals are obtained by using the easily available "equal tails" limits.

In part II we drop the restriction that the tests be based on a one or two-tailed use of the F-distribution. By a slight extension of results of Neyman and Pearson, common best critical regions for testing the hypothesis $\theta = \theta_0$ against alternatives $\theta < \theta_0$, or $\theta > \theta_0$, are found. Since the regions are always distinct for these two "one-sided" cases, there is no uniformly most powerful test. In order to find the most efficient unbiased test some recently published theorems of the writer are applied to prove that the critical region of the unbiased test proposed in part I is of type B_1.

The fact that the results summarized in the above paragraph are obtained for arbitrary positive θ_0 will immediately suggest to the reader familiar with Neyman's theory of confidence intervals that it may be easy on the basis of those results to draw conclusions about the existence of Neyman's various categories of confidence intervals. It is. In particular we find that the set of confidence intervals arrived at in §5 constitutes Neyman's short unbiased set.

The writer is aware that not all the results of this paper are new, and hopes he has given credit where it is due, but believes it desirable to bring together all the results, old and new, in this attempt to clean up the problems (*i*) and (*ii*). He is pleased to acknowledge his debt to Mr. David Votaw for aiding in the calculations for fig. 1 and for finding the formulas (6).

PART I. SIGNIFICANCE TESTS AND CONFIDENCE INTERVALS BASED ON THE F-DISTRIBUTION

2. The F-distribution The sample $E_i : (x_{i1}, x_{i2}, \cdots, x_{iN_i})$, $i = 1, 2$, is assumed to be from a normal population π_i with mean a_i and variance σ_i^2. We

write $\theta = \sigma_1^2/\sigma_2^2$, and might regard the statistic T as an estimate[1] of θ, where $T = s_1^2/s_2^2$ and

$$s_i^2 = \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2/n_i, \qquad \bar{x}_i = \sum_{j=1}^{N_i} x_{ij}/N_i, \qquad n_i = N_i - 1.$$

It will be convenient to consider θ, σ_2^2, a_1, a_2 as the population parameters, σ_1^2 being eliminated from the joint p.d.f. (probability density function) of E_1 and E_2 by the substitution $\sigma_1^2 = \theta\sigma_2^2$. For any given positive number θ_0 we define the composite hypothesis

$$H_0: \quad \theta = \theta_0, \quad 0 < \sigma_2^2 < +\infty, \quad -\infty < a_1 < +\infty, \quad -\infty < a_2 < +\infty.$$

In Hotelling's apt terminology the last three parameters are nuisance parameters.

It is well known that U_1 and U_2, where $U_i = n_i s_i^2/\sigma_i^2$, are independently distributed according to χ^2-laws with n_1 and n_2 degrees of freedom respectively, and that hence the quotient $F = (U_1/n_1) \div (U_2/n_2) = T/\theta$ has the F-distribution $h_{n_1 n_2}(F)\, dF$ with n_1 and n_2 degrees of freedom, where

$$h_{n_1 n_2}(u) = \frac{(n_1/n_2)^{\frac{1}{2}n_1}}{\mathrm{B}(\frac{1}{2}n_1, \frac{1}{2}n_2)} u^{\frac{1}{2}n_1 - 1}\left(1 + \frac{n_1}{n_2} u\right)^{-\frac{1}{2}(n_1+n_2)}, \qquad 0 \leqq u \leqq \infty.$$

For later reference we note that if we define the variable x from

$$(1) \qquad F = \frac{n_2}{n_1} \frac{x}{1 - x},$$

then the cumulative distribution function of x is the incomplete Beta function[2] $I_x(\frac{1}{2}n_1, \frac{1}{2}n_2)$.

Let α be any number such that $0 < \alpha < 1$ (α will be the significance level for (*i*); $1 - \alpha$, the confidence coefficient for (*ii*)). The symbols $A_{n_1 n_2}$, $B_{n_1 n_2}$ will always denote a pair of numbers for which[3]

$$(2) \qquad \int_{A_{n_1 n_2}}^{B_{n_1 n_2}} h_{n_1 n_2}(u)\, du = 1 - \alpha.$$

Every choice of the pair A, B leads to a solution of problems (*i*) and (*ii*):

(*i*). *A test of H_0 at significance level α consists of rejecting H_0 if $T < A_{n_1 n_2}\theta_0$ or $T > B_{n_1 n_2}\theta_0$.*

The probability of rejecting H_0 if it is true is

$$1 - Pr(A\theta_0 \leqq T \leqq B\theta_0 \mid \theta_0) = 1 - Pr(A < T/\theta_0 < B \mid \theta_0) = \alpha,$$

independently of the true values of the nuisance parameters.

---

[1] Biased.

[2] All the results of this paper pertaining to the F-distribution could of course be stated in terms of Fisher's z-distribution [2] or the incomplete Beta distribution, the first is used here because of its popularity in applied statistics, and because it permits the simplest statements for solutions of problems (*i*) and (*ii*).

[3] Superscripts on A, B will signify that a further condition has been laid on the pair A, B The subscripts will be dropped when there is no danger of confusion. We permit $B = \infty$ as a possible choice.

(ii). *A set of confidence intervals for θ with confidence coefficient $1 - \alpha$ is*[4]

$$T/B_{n_1n_2} \leq \theta \leq T/A_{n_1n_2}.$$

The probability that the true value of θ will be covered by the above random interval is

$$Pr(T/B \leq \theta \leq T/A \mid \theta) = Pr(A \leq T/\theta \leq B \mid \theta) = 1 - \alpha,$$

whatever be the true values of θ and the nuisance parameters.

It will be convenient to adopt a brief notation for the tests and confidence intervals determined by certain choices of the limits A, B. In the sequel we shall denote these choices by $A^i_{n_1n_2}$, $B^i_{n_1n_2}$, where i = I, II, $\cdots$, VI. We shall call the significance test based on the pair A^i, B^i the *test i*, and the set of confidence intervals based on this pair, the *set i of confidence intervals*, or sometimes more briefly, the *confidence intervals i*.

3. Use of one tail. Suppose a situation in which we do not mind accepting H_0 if the true value of θ exceeds θ_0, but we desire a test which is as sensitive as possible in rejecting H_0 when $\theta < \theta_0$. It can be shown (for $n_2 > 2$) that the expected value of T is $\mathcal{E}(T) = n_2\theta/(n_2 - 2)$, and hence when the true value of θ is small compared with θ_0, so is $\mathcal{E}(T)$. By the usual intuitive considerations we are led to rejecting H_0 if $F = T/\theta_0$ falls in the left tail of the F-distribution. To make the significance level equal to α we take the limits A, B so that

$$\int_0^{A^{I}_{n_1n_2}} h_{n_1n_2}(u)\,du = \alpha, \qquad B^{I}_{n_1n_2} = \infty.$$

Similarly, to test H_0 against alternatives $\theta > \theta_0$ we define test II by

$$A^{II}_{n_1n_2} = 0, \qquad \int_{B^{II}_{n_1n_2}}^{\infty} h_{n_1n_2}(u)\,du = \alpha.$$

Why test I is best for testing H_0 against alternatives $\theta < \theta_0$, and test II for $\theta > \theta_0$, will be explained more convincingly in §9.

The confidence intervals I and II are then semi-infinite. It is apparent that if we are not loath to accept large values of θ but wish to exclude the largest possible interval of small values $(0, T/B)$, we should use the set II. Indeed, the set II is optimum in the case where we are willing to accept values of θ larger than the true value but desire the highest possible probability of excluding any values less than the true value; however, the precise formulation and proof of this statement must be postponed to part II. Analogous remarks apply to the set I and a willingness to accept values of θ less than the true value.

For α = .05 or .01 the values of $B^{II}_{n_1n_2}$ are given in Snedecor's F-tables [12;

---

[4] If $B = \infty$ we omit the equality sign to the left of θ, if $A = 0$, the equality sign to the right of θ.

same n_1, n_2 as ours], and the values of $A^{\mathrm{I}}_{n_1 n_2}$ may be calculated from the same tables by using the relation

$$A^{\mathrm{I}}_{n_1 n_2} = 1/B^{\mathrm{II}}_{n_2 n_1}. \tag{3}$$

$A^{\mathrm{I}}_{n_1 n_2}$ for $\alpha = .50, .25, .10, .025, .005$ may be obtained by use of the transformation (1) and Thompson's new tables [13] of percentage points for the incomplete Beta distribution. $B^{\mathrm{II}}_{n_1 n_2}$ for these values of α can then be found from (3).

4. Symmetry condition. We now restrict our attention (until §9) to the "two-sided" situation in which we are interested in all alternatives to $\theta = \theta_0$ on the range $0 < \theta < \infty$. Let us contemplate the following *symmetry condition*:

$$A_{n_1 n_2} = 1/B_{n_2 n_1} \tag{4}$$

for all positive integers n_1, n_2. The desirability of this condition and that of §5 follows not from mathematical principles but from practical considerations which might be relevant whenever significance tests or confidence intervals are considered for a parameter θ which is the quotient of two other positive parameters θ_1 and θ_2, and the estimate of θ is the quotient of the estimates of θ_1 and θ_2.

Suppose that given the samples E_1 and E_2, computer $\mathcal{C}_1$ labels them 1, 2, the same way we have, and using our test of §2, rejects the hypothesis that $\sigma_1^2/\sigma_2^2 = k$ unless

$$A_{n_1 n_2} k \leqq s_1^2/s_2^2 \leqq B_{n_1 n_2} k;$$

while computer $\mathcal{C}_2$ labels them 2, 1, and following a similar rule rejects $\sigma_2^2/\sigma_1^2 = 1/k$ (in our notation) unless

$$A_{n_2 n_1}/k \leqq s_2^2/s_1^2 \leqq B_{n_2 n_1}/k.$$

It will be seen that (4) is merely the condition that they reach the same conclusion. This makes life simpler, at least for computers and consulting statisticians. Likewise, if $\mathcal{C}_1$ and $\mathcal{C}_2$ use the confidence intervals of §2, then they will make numerically equivalent statements about σ_1^2/σ_2^2 and σ_2^2/σ_1^2 if (4) is satisfied.

5. Logarithmically shortest confidence intervals. The length of the confidence intervals of §2 is $L = T(A^{-1} - B^{-1})$. We might consider choosing A, B in such a way that $\mathcal{E}(L)$ is minimum. This leads to the problem of minimizing $A^{-1} - B^{-1}$ subject to (2). It might seem just as desirable, however, to minimize the expected length of the confidence interval for $\theta^{\frac{1}{2}}$,

$$(T/B)^{\frac{1}{2}} \leqq \sigma_1/\sigma_2 \leqq (T/A)^{\frac{1}{2}}.$$

This leads to a different problem with a different solution.

The condition on confidence intervals for θ which appears intuitively desirable to the writer, is that the limits $\underline{\theta}$, $\bar{\theta}$ of the confidence interval $\underline{\theta}(E_1, E_2) \leqq \theta \leqq \bar{\theta}(E_1, E_2)$ be such that $\mathcal{E}(\log \bar{\theta} - \log \underline{\theta})$ is minimum. For the confidence inter-

vals of §2 this is equivalent to minimizing B/A, and by using the method of Lagrange's multipliers we easily find that

$$(5) \qquad [uh_{n_1n_2}(u)]_{u=A}^{B} = 0$$

and (2) must be satisfied. Denote the solution[5] by $A^{\mathrm{III}}_{n_1n_2}$, $B^{\mathrm{III}}_{n_1n_2}$. It is evident that the same condition (5) is obtained if we ask for logarithmically shortest confidence intervals (based on the F-distribution) for θ^k where $k > 0$.

The numerical values of the limits A^{III}, B^{III} are difficult to calculate if $n_1 \neq n_2$. The best procedure seems to be to transform to the incomplete Beta distribution by means of (1) and to calculate the corresponding points $a^{\mathrm{III}}_{n_1n_2}$, $b^{\mathrm{III}}_{n_1n_2}$ from the equations

$$(6) \qquad [I_x(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2)]_{x=a}^{b} = [I_x(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2)]_a^b = 1 - \alpha.$$

The points a, b can be found to two decimals by inspection of Pearson's tables [9]. Unfortunately, in the many cases where a is close to 0, or b to 1, A^{III}, B^{III} are then subject to enormous error when calculated from (1).

6. Reciprocal limits. While the problems (*i*) and (*ii*) are closely related, the last choice of limits was suggested solely by our consideration of (*ii*). Later we will reconsider this choice from the standpoint of (*i*),—the reader may anticipate that it will again be found advantageous in some respect. For the present, we proceed to three further choices, these arising from various approaches to (*i*).

The procedure recommended in several statistics manuals (see §8) for testing the hypothesis $\theta = 1$ is to refer the quotient of the larger of s_1^2, s_2^2 by the smaller to tables. This suggests the introduction of a statistic M defined as the maximum of T, T^{-1}. Its distribution[6] under the hypothesis $\theta = 1$ is easily found: Let $g_{n_1n_2}(M)$ be its p.d.f. Then for $1 \leqq u \leqq \infty$,

$$\begin{aligned} g_{n_1n_2}(u)\,du &= Pr(u < M < u + du \mid \theta = 1) \\ &= Pr(u < T < u + du \text{ or } u < T^{-1} < u + du) \\ &= Pr(u < T < u + du) + Pr(u < T^{-1} < u + du), \end{aligned}$$

since the last two terms are the probabilities of mutually exclusive events. Furthermore, the first term is $h_{n_1n_2}(u)\,du$, and because of the symmetry induced by $\theta_0 = 1$ we can evaluate the second term by merely interchanging subscripts. Hence the desired distribution is

$$g_{n_1n_2}(u) = h_{n_1n_2}(u) + h_{n_2n_1}(u),$$

regardless of the true values of the nuisance parameters.

[5] It can be shown by elementary methods that the solution of these equations exists and is unique; likewise for the solutions later denoted by superscripts IV and V.

[6] Considered by K. Pearson [8].

If we reject the hypothesis $\theta = 1$ if $M > M_{n_1 n_2}$, where

$$\int_{M_{n_1 n_2}}^{\infty} g_{n_1 n_2}(u)\, du = \alpha,$$

then this significance test is easily shown to be the same as that of §2 with $\theta_0 = 1$ and

$$A_{n_1 n_2} = B_{n_1 n_2}^{-1}.$$

We remark that again these limits are not easy to compute if $n_1 \neq n_2$. While this choice of A, B, which we shall call $A_{n_1 n_2}^{\mathrm{IV}}$, $B_{n_1 n_2}^{\mathrm{IV}}$, has been motivated only for the case $\theta_0 = 1$, it leads of course to a test IV for any θ_0 and a set IV of confidence intervals.

7. The likelihood ratio. Since the properties of λ-criteria in general have received much attention in the literature, and since in particular the λ-test for H_0 is equivalent to a certain choice of A, B, we shall mention it here, and see whether it has any advantages in §9. λ for H_0 in the case $\theta_0 = 1$ was given by Pearson and Neyman [7; their H_1, n_i, s_i^2, θ, λ_{H_1} are our H_0, N_i, $s_i^2(N_i - 1)/N_i$, $N_1(N_2 - 1)/\{N_2(N_1 - 1)T\}$, λ]; for any θ_0 it may be shown to be

$$\lambda = C_{n_1 n_2} F^{3/2} \left(1 + \frac{n_1}{n_2} F\right)^{-1} h_{n_1 n_2}(F).$$

On considering the (bell-shaped) graph of λ against F we see that $\lambda < \lambda_0$ corresponds to two intervals, say $0 \leqq F < F'$ and $F'' < F \leqq \infty$. The λ-test, which consists of rejecting H_0 when $\lambda < \lambda_0$, where λ_0 is determined so that the significance level is α, is thus equivalent to test V with $A_{n_1 n_2}^{\mathrm{V}}$, $B_{n_1 n_2}^{\mathrm{V}}$ satisfying (2) and

$$\left[u^{3/2} \left(1 + \frac{n_1}{n_2} u\right)^{-1} h_{n_1 n_2}(u) \right]_{u=A}^{B} = 0.$$

8. Equal tails. Perhaps the most venerable procedure for determining limits on a distribution for a significance test in a "two-sided" case is to choose them so that the tails of the distribution have equal areas. Define $A_{n_1 n_2}^{\mathrm{VI}}$, $B_{n_1 n_2}^{\mathrm{VI}}$ from

$$\int_0^{A_{n_1 n_2}^{\mathrm{VI}}} h_{n_1 n_2}(u)\, du = \int_{B_{n_1 n_2}^{\mathrm{VI}}}^{\infty} h_{n_1 n_2}(u)\, du = \tfrac{1}{2}\alpha$$

The values of $B_{n_1 n_2}^{\mathrm{VI}}$ for $\alpha = .10$ and $.02$ are given in the F-tables [12; same n_1, n_2 as ours] as 5% and 1% points. The relation

$$A_{n_1 n_2}^{\mathrm{VI}} B_{n_2 n_1}^{\mathrm{VI}} = 1 \tag{7}$$

is easy to get, and hence $A_{n_1 n_2}^{\mathrm{VI}}$ for these values of α may also be calculated from the F-tables. The limits for $\frac{1}{2}\alpha = .25, .10, .025, .005$ can be calculated from (1), (7), and Thompson's tables [13]

Since test VI will later be seen to have some merit we will discuss it somewhat further at this point. In several statistics texts [e.g., 3, 14] the student is told to take the quotient of the larger by the smaller of s_1^2, s_2^2, refer it to the F-table, taking the n_1 of the table to be the n, of the numerator, and to reject the null hypothesis $\theta = 1$ if the sample value is larger than the tabulated. It is then further stated without proof that in using the 5% or 1% points of the F-table, the significance level is actually 10% or 2%. Since the quotient thus referred to the table is precisely the statistic M of §6, it would seem logical to refer it to an M-table rather than the F-table! However, the above procedure can be justified[7] as follows: The equation (7) tells us that test VI fulfills the symmetry condition (4). It makes no difference then in his conclusions whether the computer uses the statistic s_1^2/s_2^2 and the distribution $h_{n_1n_2}(F)$ or s_2^2/s_1^2 and $h_{n_2n_1}(F)$. In particular he may always use the larger ratio and $h_{mn}(F)$, where m and n are the "degrees of freedom" of numerator and denominator, respectively. Since this statistic cannot fall in the lower tail, he need consider only whether the calculated value exceeds the tabulated. *But in using the value tabulated as the upper p% point of the F-distribution, he makes his test at the 2p% significance level.*

9. Comparison of the tests and confidence intervals. We now have at hand two one-tailed and four two-tailed tests, and corresponding sets of confidence intervals, all based on the F-distribution. We note at this point that all four of the two-tailed tests satisfy the symmetry condition (4), and that in the special case $n_1 = n_2$, these four tests become identical. In comparing any two tests, an instrument which makes their relative advantages completely *anschaulich* is the power curve (surface in a more complicated case). The definition and interpretation of the power curve of a test are based on the insight of Neyman and Pearson [5] that two types of error are possible in applying a test: We may (I) reject the hypothesis when it is true, or (II) accept it when it is false.

We see immediately that for any test of the class considered in §2, the probability of a type I error is the same, namely α. To find the probability of a type II error, let us introduce a little more terminology: We denote by E the sample point (E_1, E_2) and by w the region of sample space defined by

$$T < A\theta_0 \quad \text{and} \quad T > B\theta_0 . \tag{8}$$

w is called the *critical region* of the test: the test rejects H_0 if and only if E falls in w. The probability of this, which is called the *power* of the test, is

$$1 - Pr(A\theta_0/\theta \leqq T/\theta \leqq B\theta_0/\theta \mid \theta, \sigma_2^2, a_1, a_2).$$

Since in the present case this happens to be completely independent of the true values of the nuisance parameters, even for $\theta \neq \theta_0$, let us write it as $P(w \mid \theta)$. Then

---

[7] The writer is indebted to Mr T. W. Anderson, Jr. for pointing out to him that it is not necessary to use the M-distribution.

(9) $$P(w \mid \theta) = 1 - \int_{A\theta_0/\theta}^{B\theta_0/\theta} h_{n_1 n_2}(u)\, du.$$

Finally, by the *power curve* of the test we mean simply the graph of the power $P(w \mid \theta)$ as a function of θ

We may now state the probability of a type II error: it is $1 - P(w \mid \theta)$, where necessarily $\theta \neq \theta_0$ Hence the ordinate on the power curve for $\theta \neq \theta_0$ is the probability of avoiding a type II error, while for $\theta = \theta_0$ it is the probability of making a type I error. By inspection of equation (9) we find that, barring the cases $B = \infty$ or $A = 0$ (tests I and II), $P(w \mid \theta) \to 1$ as $\theta \to 0$ or ∞. We calculate the derivative to be

(10) $$P'(w \mid \theta) = [uh_{n_1 n_2}(u)/\theta]_{u=A\theta_0/\theta}^{B\theta_0/\theta},$$

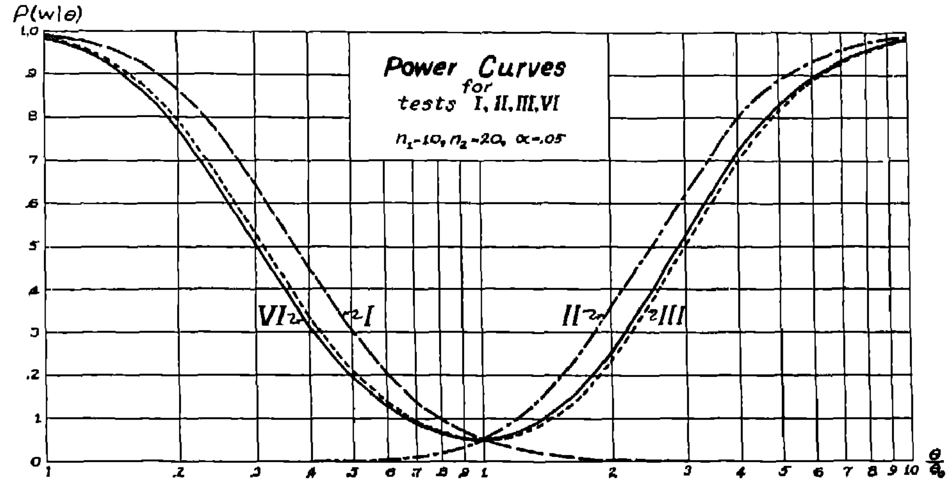


FIG 1

which is obviously continuous for $0 < \theta < \infty$. If we equate this to zero we find a unique solution for θ, and hence the power curve has a single minimum point. In the exceptional case $B = \infty$ we see from (9) that $P(w \mid \theta)$ decreases monotonically from 1 to 0 as θ increases from 0 to ∞; in the case $A = 0$, $P(w \mid \theta)$ increases monotonically from 0 to 1. Some power curves[8] are plotted in fig. 1.

Always understanding by w a region of the set defined by (8), and recalling the above interpretation of the ordinate on the power curve, we are led to ask whether there is not a w, say w_0, whose power curve nowhere drops below any other curve $P = P(w \mid \theta)$. (They all pass through (θ_0, α).) The test based on such a region w_0 would be called *uniformly most powerful* (UMP) of the class considered, and obviously would be preferred under any circumstances. Alas,

[8] Power curves for test V may be found in a paper by Brown [1]. It did not seem worthwhile to construct curves for test IV, since the limits are hard to compute, the test is biased, and has little historical interest

it does not exist Perhaps some insight into the fact of the general non-existence of UMP tests can be gained by returning to fig. 1. While fig. 1 is for the case $n_1 = 10$, $n_2 = 20$, and $\alpha = .05$, the following remarks are valid for any n_1, n_2, α: We note that for testing H_0 against alternatives $\theta < \theta_0$ test I is far superior to the other three, indeed it is superior to any of the tests of the class defined by (8) in the sense that its power curve lies above that of any of the other tests.[9] But for alternatives $\theta > \theta_0$, test I is seen to be very poor (the worst possible, it can be shown). Similar remarks apply to test II and the complementary alternatives. This constitutes the more convincing explanation promised in §2 of the superiority of tests I and II in the "one-sided" cases. Since the power curve of test I lies above all other power curves for $\theta < \theta_0$, and that of test II above all for $\theta > \theta_0$, it is now clear that there is no UMP test of the class considered.

To cope with the commonly occurring situation where there is no UMP test, Neyman and Pearson [5] defined an *unbiased* test,—one whose power curve has an absolute minimum at θ_0. The desirability of an unbiased test in the "two-sided" case is evident when we note that if a test is biased, the probability that we accept the hypothesis $\theta = \theta_0$ is greater if θ has certain values $\theta \neq \theta_0$ than if $\theta = \theta_0$. To find which, if any, of our tests is unbiased, we equate expression (10) to zero for $\theta = \theta_0$. As a result we find[10] the condition (5) which determines test III.

We see now that the limits A^{III}, B^{III} yield the preferred test in the "two-sided" case, as well as the logarithmically shortest confidence intervals. However, as pointed out in §5, the numerical values of these limits are difficult to calculate, and the question then arises, do we lose much by using instead the easily obtained "equal tails" limits A^{VI}, B^{VI}? In the case $n_1 = 10$, $n_2 = 20$, $\alpha = .05$, fig. 1 shows that the power curves of tests III and VI differ very little. The extent of the bias of test VI for other values of n_1, n_2, and $\alpha = .05$, .01 is indicated in table I. (The missing diagonal entries are all 1,5 or 1,1). Let us call the entries β, 100 $\bar{\alpha}$, where $\beta = \theta_{\min}/\theta_0$, $\bar{\alpha} = P(w^{\mathrm{VI}} \mid \theta_{\min})$. From (10) and (1) we get the following formula for computing β:

$$\beta = (\mathcal{B} - \mathcal{A}Q^{n_1/(n_1+n_2)})/(Q - 1),$$

where

$$Q = \mathcal{B}/\mathcal{A}, \qquad \mathcal{A} = a/(1 - a), \qquad \mathcal{B} = b/(1 - b),$$

and a and $1 - b$ are the $100(\frac{1}{2}\alpha)\%$ points on the incomplete Beta distribution for $\nu_2 = n_1$, $\nu_1 = n_2$, and $\nu_1 = n_1$, $\nu_2 = n_2$, respectively, in the notation of Thompson's tables [13]. $\bar{\alpha}$ may then be computed by transforming (9),

$$\bar{\alpha} = 1 - \Big[I_x(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) \Big]_{x = (1 + \beta/\mathcal{A})^{-1}}^{(1 + \beta/\mathcal{B})^{-1}},$$

[9] The reader may prove this from (9) or note that it is a special case of the results of §10.

[10] The equivalent condition on the incomplete Beta distribution was given by Pitman [10] for the case $\theta_0 = 1$.

TABLE I

Minimum points of power curves of test VI

The entries are $\theta_{\min}/\theta_0$, $100\,P(w^{VI} \mid \theta_{\min})$,

Roman type for $\alpha =$ 05, **bold face for $\alpha =$ .01**

n_1 \ n_2	1	2	3	5	10	20	40	∞
1		634, 4.75 **.631, .946**	576, 4 47 **.577, .883**	.559, 4.17 **.571, .808**	.565, 3.89 **.595, .740**	.574, 3.75 **.617, .705**	581, 3 68 **.630, .687**	588, 3.61 **.645, .670**
2	1.578, 4 75 **1.585, .946**		861, 4.93 **.855, .982**	.779, 4.69 **.776, .928**	.745, 4.44 **.749, .853**	737, 4 26 **.749, .804**	.735, 4 15 **.753, .778**	.735, 4.05 **.760, .751**
3	1 735, 4.47 **1.734, .883**	1 161, 4.93 **1.170, .982**		.895, 4 92 **.889, .978**	.838, 4.70 **.835, .917**	.819, 4 51 **.821, .867**	.812, 4.41 **.819, .837**	.808, 4.29 **.820, .804**
5	1.789, 4 17 **1.752, .808**	1 284, 4 69 **1.289, .928**	1.117, 4 92 **1.124, .978**		927, 4.92 **.924, .975**	.898, 4.78 **.896, .934**	886, 4.67 **.887, .903**	.877, 4.54 **.882, .864**
10	1.771, 3.89 **1.682, .740**	1.342, 4 44 **1.335, .853**	1.194, 4.70 **1.198, .917**	1.079, 4.92 **1.083, .975**		.965, 4.96 **.964, .987**	.949, 4.89 **.949, .964**	.941, 4.76 **.937, .925**
20	1.742, 3.75 **1.622, .705**	1.357, 4.26 **1.335, .804**	1.221, 4 51 **1.217, .867**	1 114, 4.78 **1.116, .934**	1.036, 4.96 **1.038, .987**		.983, 4.98 **.983, .993**	.967, 4.88 **.968, .960**
40	1.722, 3.68 **1.587, .687**	1.360, 4.15 **1.327, .778**	1 231, 4.41 **1.221, .837**	1.129, 4.67 **1.127, .903**	1 053, 4.89 **1.054, .964**	1.017, 4.98 **1.018, .993**		.984, 4.94 **.984, .980**
∞	1 700, 3.61 **1.549, .670**	1.360, 4 05 **1.316, .751**	1.238, 4.29 **1.219, .804**	1.140, 4.54 **1.134, .864**	1.063, 4.76 **1.067, .925**	1.034, 4.88 **1.034, .960**	1.017, 4.94 **1.017, .980**	

and using Pearson's tables [9], or, when x is very close to 0 or 1, using a few terms of the series

$$I_\delta(\tfrac{1}{2}m, \tfrac{1}{2}n) = 1 - I_{1-\delta}(\tfrac{1}{2}n, \tfrac{1}{2}m) = \frac{\delta^{\frac{1}{2}m}}{B(\frac{1}{2}m, \frac{1}{2}n)}\left[\frac{2}{m} - \frac{n-2}{2^0(m+2)}\frac{\delta}{1!} + \frac{(n-2)(n-4)}{2^1(m+4)}\frac{\delta^2}{2!} - \frac{(n-2)(n-4)(n-6)}{2^2(m+6)}\frac{\delta^3}{3!} + \cdots\right].$$

In computing β, $\bar{\alpha}$ it is perhaps simplest to take $n_1 > n_2$ and use the relationships

$$\beta_{n_1 n_2} = 1/\beta_{n_2 n_1}, \qquad \bar{\alpha}_{n_1 n_2} = \bar{\alpha}_{n_2 n_1} .$$

When sample sizes $n_1 + 1$, $n_2 + 1$ are such that table I indicates a large bias' it might be worthwhile to get limits for an unbiased test from the "equal tails', limits as follows: The limits $\bar{A}^{\mathrm{III}}$, $\bar{B}^{\mathrm{III}}$ for an unbiased test III may be obtained by taking

$$\bar{A}^{\mathrm{III}} = A^{\mathrm{VI}}/\beta, \qquad \bar{B}^{\mathrm{III}} = B^{\mathrm{VI}}/\beta,$$

but *the test will then be at significance level* $\bar{\alpha}$. The gain in using $\bar{A}^{\mathrm{III}}$, $\bar{B}^{\mathrm{III}}$ instead of A^{VI}, B^{VI} is more apparent when we consider confidence intervals: The sets associated with $\bar{A}^{\mathrm{III}}$, $\bar{B}^{\mathrm{III}}$, and A^{VI}, B^{VI} have the same logarithmic lengths, but the confidence coefficients are $1 - \bar{\alpha}$ and $1 - \alpha$, respectively.

This seems to be about as far as it is worthwhile to carry the developments at the elementary level of part I. Some inadequacies may already have disturbed the reader: Why not consider in place of the interval (A, B) on the range of F any measurable region[11] R such that the integral of $h_{n_1 n_2}(F)$ over R is $1 - \alpha$? Under the transformation $T = \theta_0 F$ the complement of R, just as the complement of (A, B), would lead to critical regions w for which $P(w \mid \theta_0) = \alpha$ for all values of the nuisance parameters. Critical regions satisfying the last condition are said to be *similar* to the sample space with regard to the nuisance parameters. More generally, how would our preferred test I, II, III stand up if we admit for comparison, tests based on any similar regions whatever? Finally, how can one formulate in a general way conditions for optimum confidence intervals, and would a more general formulation still lead to the preference of the sets I, II, III? Answers to these questions will be found in part II.

Part II. Significance Tests and Confidence Intervals Based on any Similar Regions

10. Common best critical regions. For the case $\theta_0 = 1$, Neyman and Pearson [6] have shown that the critical region of test I is the common best critical (CBC) region for testing H_0 against alternatives $\theta < \theta_0$. This result is easily extended to any θ_0 by a simple device. We consider the following 1:1 transformations of variables and parameters:

---

[11] Our intuitions may balk at the notion of using sets R more general than intervals, but it would nevertheless be reassuring to find that our tests can meet this competition.

$$(11) \qquad x_{1j} = \theta_0^{\frac{1}{2}} x_{1j}', \qquad x_{2k} = x_{2k}', \qquad j = 1, 2, \cdots, N_1; k = 1, 2, \cdots, N_2,$$

$$(12) \qquad \theta = \theta_0\theta', \qquad \sigma_2^2 = (\sigma_2')^2, \qquad a_1 = \theta_0^{\frac{1}{2}} a_1', \qquad a_2 = a_2'.$$

Denote by E_1', E_2', E' the points corresponding to E_1, E_2, E, respectively, under the transformation (11), by ϑ any point in the space of the three nuisance parameters, and by ϑ' its correspondent under the transformation (12), by H_0' the transformed hypothesis, H_0'. $\theta' = 1$, ϑ', unspecified. If w is any Borel-measurable region of the space of E, and w' the map of w under (11), then $Pr(E \epsilon w \mid \theta, \vartheta) = Pr(E' \epsilon w' \mid \theta', \vartheta')$, which we shall write as

$$(13) \qquad P(w \mid \theta, \vartheta) = P(w' \mid \theta', \vartheta').$$

We note that the coordinates of E_i' are normally distributed with mean a_i' and variance $(\sigma_i')^2$ where $(\sigma_1')^2 = \theta'(\sigma_2')^2$, all $N_1 + N_2$ coordinates being statistically independent. Designating the critical region of test I by w_0, and its map under (11) by w_0', the result of Neyman and Pearson may then be stated as follows. w_0' is a CBC region for H_0' and alternatives $\theta' < 1$. Now suppose w_0 were not a CBC region for H_0 and alternatives $\theta < \theta_0$. Then there would exist a region w_1, a value $\theta_1 < \theta_0$, and a point ϑ_1 such that $P(w_1 \mid \theta_1, \vartheta_1) > P(w_0 \mid \theta_1, \vartheta_1)$, while $P(w_1 \mid \theta_0, \vartheta) = \alpha$ for all ϑ. Let w_1', θ_1', ϑ_1' correspond to w_1, θ_1, ϑ_1 under (11) and (12). Then from (13) we would have that $P(w_1' \mid \theta_1', \vartheta_1') > P(w_0' \mid \theta_1', \vartheta_1')$, where $\theta_1' < 1$, while $P(w_1' \mid 1, \vartheta') = \alpha$ for all ϑ'. But this would contradict the fact that w_0' is a CBC region for H_0' and alternatives $\theta' < 1$.

The proof that the critical region of test II is a CBC region for testing H_0 against alternatives $\theta > \theta_0$ is of course completely analogous. This establishes the non-existence of a UMP test for H_0, and so we consider next the existence of a "best" unbiased test.

11. Type B_1 region. This section is a direct application of a recent paper "On the theory of testing composite hypotheses with one constraint" to which we shall refer as [11] Since it is not feasible to restate here the definitions, assumptions, and theorems of [11], we shall refer to them by their numbers there It is convenient to transform the parameters of the p.d.f. of E by putting

$$(14) \qquad \theta = 1/\psi, \qquad \theta_0 = 1/\psi_0, \qquad \sigma_2^2 = 1/h.$$

Then

$$(15) \quad p(E \mid \psi, h, a_1, a_2) = (2\pi)^{-\frac{1}{2}N} \psi^{\frac{1}{2}N_1} h^{\frac{1}{2}N} \cdot \exp\{-\tfrac{1}{2}\psi h[N_1(\bar{x}_1 - a_1)^2 + S_1] + h[N_2(\bar{x}_2 - a_2)^2 + S_2]\},$$

where

$$N = N_1 + N_2, \qquad S_i = n_i s_i^2.$$

We note that type B and type B_1 regions (definitions 1, 2 in [11]) are invariant under certain transformations of parameters: Suppose new parameters θ', ϑ'

are introduced by 1:1 transformations $\theta = \theta(\theta')$, $\vartheta = \vartheta(\vartheta')$. Let θ_0' correspond to θ_0, and consider the transformed hypothesis $H_0' : \theta' = \theta_0'$; ϑ', unspecified. Sufficient conditions that a region be of type B for testing H_0 if it is of type B for testing H_0' are that the function $\theta(\theta')$ have first and second derivatives and that the first not vanish at θ_0'. The last statement remains true if B is replaced by B_1. Since the transformations (14) satisfy these sufficient conditions, we define

$$H_0' : \quad \psi = \psi_0 ; \quad \vartheta' = (h, a_1, a_2), \text{ unspecified,}$$

and propose to show that there exists a type B_1 region for testing H_0', and that it is the critical region of test III.

For later reference we now note that the four functions of variables and parameters defined in Table II are mutually independently distributed as indicated there.

TABLE II

Function	Distribution
$U_1 = \psi h S_1 = S_1/\sigma_1^2$	χ^2, with n_1 degrees of freedom
$U_2 = hS_2 = S_2/\sigma_2^2$	" " n_2 " " "
$u_3 = (\psi h N_1)^{\frac{1}{2}}(\bar{x}_1 - a_1) = N_1^{\frac{1}{2}}(\bar{x}_1 - a_1)/\sigma_1$	normal, with zero mean and unit variance
$u_4 = (hN_2)^{\frac{1}{2}}(\bar{x}_2 - a_2) = N_2^{\frac{1}{2}}(\bar{x}_2 - a_2)/\sigma_2$	" " " " " " "

Let us first verify the critical assumption 3^0 of [11]: Identifying our ψ, h, a_1, a_2 with $\theta_1, \theta_2, \theta_3, \theta_4$ of [11], we find from (15) that

$$\begin{aligned}
\phi_1 &= \tfrac{1}{2}\{N_1/\psi - h[N_1(\bar{x}_1 - a_1)^2 + S_1]\},\\
\phi_2 &= \tfrac{1}{2}\{N/h - \psi[N_1(\bar{x}_1 - a_1)^2 + S_1] - [N_2(\bar{x}_2 - a_2)^2 + S_2]\},\\
\phi_3 &= \psi h N_1(\bar{x}_1 - a_1),\\
\phi_4 &= hN_2(\bar{x}_2 - a_2),
\end{aligned} \tag{16}$$

and then check 3^0 by differentiating equations (16).

To verify assumption 4^0, let x_1, x_2, x_3, x_4 of [11] be our $x_{11}, x_{12}, x_{21}, x_{22}$, respectively. We calculate

$$\frac{\partial(\phi_1, \phi_2, \phi_3, \phi_4)}{\partial(x_1, x_2, x_3, x_4)} = \psi h^3(x_1 - x_2)(x_4 - x_3),$$

which vanishes only on the same set of probability zero for all admissible values of the parameters. The validity of assumption 5^0 follows from §5 of [11], and there is no difficulty in verifying 1^0 and 2^0.

To apply theorem 1 of [11] we must find functions $k_i(\phi_2, \phi_3, \phi_4 ; \psi_0, \vartheta')$, $i = 1, 2$, such that

$$\int_{k_1}^{k_2} \phi_1^t p(\phi_1, \phi_2, \phi_3, \phi_4 \mid \psi_0, \vartheta')\, d\phi_1 = (1 - \alpha) \int_{-\infty}^{+\infty} \text{same}, \tag{17}$$

for $t = 0, 1$, where the *symbols* ϕ_i *henceforth are understood to stand for the functions* (16) *with* ψ *replaced by* ψ_0. If the functions k_i exist, then the region in sample space defined by

(18) $$\phi_1 < k_1 \quad \text{and} \quad \phi_1 > k_2$$

is independent of ϑ' and of type B

From equations (16) and Table II we see that

(19) $$\begin{aligned} \phi_1 &= \tfrac{1}{2}(N_1 - u_1)/\psi_0, & \phi_2 &= \tfrac{1}{2}(N - u_2)/h, \\ \phi_3 &= (\psi_0 h N_1)^{\frac{1}{2}} u_3, & \phi_4 &= (hN_2)^{\frac{1}{2}} u_4, \end{aligned}$$

where

$$u_1 = U_1 + u_3^2, \qquad u_2 = U_1 + U_2 + u_3^2 + u_4'^2,$$

and ψ is put equal to ψ_0 in U_1, u_3. Furthermore, for fixed u_2, u_3, u_4, the range of u_1 is

$$u_3^2 \leqq u_1 \leqq u_2 - u_4^2.$$

Transforming the integrals in (17) by substituting (19) and

$$p(\phi_1, \phi_2, \phi_3, \phi_4 \mid \psi_0, \vartheta') = \frac{p(U_1, U_2, u_3, u_4 \mid \psi_0, \vartheta')}{\dfrac{\partial(\phi_1, \phi_2, \phi_3, \phi_4)}{\partial(u_1, u_2, u_3, u_4)} \cdot \dfrac{\partial(u_1, u_2, u_3, u_4)}{\partial(U_1, U_2, u_3, u_4)}},$$

where the p.d.f. in the numerator is, from Table II,

$$CU_1^{\frac{1}{2}n_1-1} U_2^{\frac{1}{2}n_2-1} \exp(-\tfrac{1}{2}u_2),$$

we get as the equivalent of (17)

$$\int_{K_1}^{K_2} (N_1 - u_1)^t (u_1 - u_3^2)^{\frac{1}{2}n_1-1} (u_2 - u_4^2 - u_1)^{\frac{1}{2}n_2-1}\, du_1 = (1 - \alpha) \int_0^1 \text{same}$$

with

$$K_i(u_2, u_3, u_4; \psi_0, \vartheta') = k_i(\phi_2, \phi_3, \phi_4; \psi_0, \vartheta').$$

Finally, we let

(20) $$x = (u_1 - u_3^2)/(u_2 - u_3^2 - u_4^2),$$

and get

$$\int_{\kappa_1}^{\kappa_2} [N_1 - u_3^2 - (u_2 - u_3^2 - u_4^2)x]^t x^{\frac{1}{2}n_1-1} (1 - x)^{\frac{1}{2}n_2-1}\, dx = (1 - \alpha) \int_0^1 \text{same},$$

where $\kappa_i(u_2, u_3, u_4; \psi_0, \vartheta')$ are the values of x obtained by setting u_1 equal to the function K_i in (20). The last condition is equivalent to

(21) $$\int_{\kappa_1}^{\kappa_2} x^{\frac{1}{2}n_1-1+t} (1 - x)^{\frac{1}{2}n_2-1}\, dx = (1 - \alpha) \int_0^1 \text{same}, \qquad t = 0, 1.$$

Since x is a continuous monotonic function of ϕ_1, (18) becomes

$$x < \kappa_1 \quad \text{and} \quad x > \kappa_2 . \tag{22}$$

Solutions for the functions κ_1, κ_2 satisfying (21) exist in the form κ_i = constant. Indeed, if we now note that the x defined by (20) is the same as that defined in (1), and let $\kappa_1 = a$, $\kappa_2 = b$, we see that the conditions (21) are identical with (6), and that our method of finding type B regions has led us to the critical region of test III.

To show that the type B region obtained from Theorem 1 of [11] is also of type B_1, we appeal to Theorem 2: From (15) we have

$$p(E \mid \psi, \vartheta')/p(E \mid \psi_0, \vartheta') = (\psi/\psi_0)^{\frac{1}{2}N_1} \exp \{(\psi - \psi_0)(\phi_1 - \tfrac{1}{2}N_1/\psi_0)\}.$$

Since for $\psi \neq \psi_0$ this function is convex in ϕ_1, Theorem 2 is applicable. The result of this section is the conclusion that the critical region of test III is of type B_1 for testing H_0.

12. Neyman's categories of confidence intervals. The concepts and terminology of this section are those formulated in a basic paper [4] by Neyman. Suppose a distribution depends on a parameter θ, and on further parameters $\theta_2, \theta_3, \cdots, \theta_l$ which we shall symbolize by ϑ. The hypothesis

$$H(\theta_0): \qquad \theta = \theta_0 ; \qquad \vartheta, \text{ unspecified},$$

may be called a composite hypothesis with one constraint [11]. Let E be the sample point, W be the sample space, and w be any Borel-measurable region in W. Write $Pr\{E \epsilon w \mid \theta, \vartheta\} = P\{w \mid \theta, \vartheta\}$. The condition that a critical region $w(\theta_0)$ for testing $H(\theta_0)$ be similar to W with respect to ϑ is

$$P\{w(\theta_0) \mid \theta_0, \vartheta\} = \alpha \text{ for all } \vartheta, \tag{23}$$

where α is fixed throughout our discussion Suppose for every admissible θ_0 there exists a similar region $w(\theta_0)$. The complementary region $A(\theta_0) = W - w(\theta_0)$ we may call a region of acceptance. For any E we next define the linear set $\delta(E)$ of points on the θ-axis as the totality of points θ such that $E \epsilon A(\theta)$. The probability [4] that the random set $\delta(E)$ cover a value θ'' if the true value of θ is θ' is

$$Pr\{\theta'' \epsilon \delta(E) \mid \theta', \vartheta\} = 1 - P\{w(\theta'') \mid \theta', \vartheta\}, \tag{24}$$

and hence from (23),

$$Pr\{\theta' \epsilon \delta(E) \mid \theta', \vartheta\} = 1 - \alpha \tag{25}$$

for all θ', ϑ, and we might call the aggregate $\{\delta(E)\}$ a set of confidence regions with confidence coefficient $1 - \alpha$. Now if all $\delta(E)$ are intervals, then they form a set of confidence intervals.

We have now shown that if $H(\theta_0)$ is a composite hypothesis with one constraint, if for every admissible θ_0 there exists a similar region $w(\theta_0)$ for testing

$H(\theta_0)$, and if the aggregate $\{\delta(E)\}$ determined by the family $\{w(\theta_0)\}$ consists of *intervals* $\delta(E)$, then $\{\delta(E)\}$ is a set of confidence intervals. By similar use of (24) the reader may prove that if furthermore each $w(\theta_0)$ of the family has the property P of the table below, then the corresponding set $\{\delta(E)\}$ of confidence intervals is of Neyman's category C.

P property of $w(\theta_0)$	C: category of $\{\delta(E)\}$
gives UMP test	shortest
CBC for $\theta > \theta_0$ (or $\theta < \theta_0$)	best one-sided
gives unbiased test	unbiased
of type B	short unbiased
of type B_1	shortest unbiased

We have taken the liberty of calling a set of one-sided confidence intervals

$$\delta(E). \qquad \underline{\theta}(E) \leqq \theta \text{ (or } \theta \leqq \bar{\theta}(E)),$$

where $\underline{\theta}(E)$ and $\bar{\theta}(E)$ are Neyman's *unique lower and upper estimates*, respectively, *best one-sided*, and of calling a set $\{\delta_0(E)\}$ *shortest unbiased* if for all θ', ϑ it satisfies (25) and

$$[\partial Pr\{\theta' \in \delta_0(E) \mid \theta, \vartheta\}/\partial\theta]_{\theta=\theta'} = 0, \tag{26}$$

while for any other set $\{\delta_1(E)\}$ satisfying (25) and (26), and all θ'', θ', ϑ,

$$Pr\{\theta'' \in \delta_0(E) \mid \theta', \vartheta\} \leqq Pr\{\theta'' \in \delta_1(E) \mid \theta', \vartheta\}.$$

It follows immediately from this discussion that our sets II and I of confidence intervals are the best one-sided, and that the set III is not only a short, but the shortest, unbiased set.

In conclusion, we remark that Neyman's concept of the "shortness" of a set of confidence intervals strikes one at first as indirect,—to fully appreciate its elegance it is perhaps necessary to attempt the formulation of a general theory from a more naive approach,—and that it is then of interest to discover that in the present case his short unbiased set coincides with that reached by the direct intuitive (but obviously extremely limited) method of §5.

REFERENCES

[1] G. W Brown, "On the power of the L_1 test for equality of several variances," *Annals of Math. Stat.*, Vol. 10 (1939), p 127
[2] R. A. Fisher, "On a distribution yielding the error function of several well known statistics," *Proc. Int. Math. Congress, Toronto, 1924*, Vol. 2, p. 808.
[3] J F Kenney, *Mathematics of Statistics*, part 2, N. Y., 1939, p. 144.
[4] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil Trans Roy Soc London*, ser. A, Vol. 236 (1937), pp 333–380
[5] J. Neyman and E S. Pearson, "Contributions to the theory of testing statistical hypotheses, part I," *Stat Res. Mem.*, Vol. 1 (1936), pp. 1–37.

[6] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy Soc. London*, ser A, Vol. 231 (1933), pp. 289–337.
[7] E. S. PEARSON and J. NEYMAN, "On the problem of two samples," *Bull. Int. Acad. Polon. Sc. Let.*, ser. A, 1930, p 82.
[8] K. PEARSON, S. A. STOUFFER, and F. N. DAVID, "Further applications in statistics of the $T_m(x)$ Bessel function," *Biometrika*, Vol. 24 (1932), pp. 306, 339, 340.
[9] K. PEARSON (Editor), *Tables of the Incomplete Beta Function*, Cambridge, 1934.
[10] E. J. G. PITMAN, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), p. 207.
[11] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 280–293.
[12] G. W. SNEDECOR, *Statistical Methods*, Ames, 1940, pp. 184–187.
[13] C. M. THOMPSON, "Tables of percentage points of the incomplete Beta function," *Biometrika*, Vol. 32 (1941), pp. 168–181.
[14] L. H. C. TIPPETT, *The Methods of Statistics*, London, 1937, p. 118.

SETTING OF TOLERANCE LIMITS WHEN THE SAMPLE IS LARGE

By Abraham Wald

Columbia University

1. Introduction. Let $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ be the joint probability density function of the variates $x_1, \cdots, x_p$ involving k unknown parameters $\theta_1, \cdots, \theta_k$. A sample of size n is drawn from this population. Denote by $x_{i\alpha}(i = 1, \cdots, p; \alpha = 1, \cdots, n)$ the α-th observation on x_i. We will deal here with the following two problems of setting tolerance limits, which are of importance in the mass production of a product:

Problem 1. *For any two positive numbers $\beta < 1$ and $\gamma < 1$ we have to construct p pairs of functions of the observations $L_i(x_{11}, \cdots, x_{pn})$ and $U_i(x_{11}, \cdots, x_{pn})$ $(i = 1, \cdots, p)$ such that*

$$(1) \quad P\left\{\int_{L_p}^{U_p} \cdots \int_{L_1}^{U_1} f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)\, dx_1 \cdots dx_p \geq \gamma \mid \theta_1, \cdots, \theta_k\right\} = \beta,$$

where for any relation R, $P(R \mid \theta_1, \cdots, \theta_k)$ denotes the probability that R holds, calculated under the assumption that $\theta_1, \cdots, \theta_k$ are the true values of the parameters.

Problem 2. *For any positive numbers $\beta < 1$, $\lambda < 1$ and for any positive integer N we have to construct p pairs of functions of the observations $L_i(x_{11}, \cdots, x_{pn})$ and $U_i(x_{11}, \cdots, x_{pn})$ with the following property: Let $y_{i\alpha}(i = 1, \cdots, p;\ \alpha = 1, \cdots, N)$ be the α-th observation on the variate x_i in a second sample of size N drawn from the same population as the first sample has been drawn. Denote by M the number of different values of α for which the p inequalities*

$$L_i(x_{11}, \cdots, x_{pn}) \leq y_{i\alpha} \leq U_i(x_{11}, \cdots, x_{pn}) \quad (i = 1, \cdots, p),$$

are fulfilled. Then

$$(2) \qquad P(M \geq \lambda N \mid \theta_1, \cdots, \theta_k) = \beta,$$

where $\theta_1, \cdots, \theta_k$ denote the unknown parameter values of the population from which the observations $x_{i\alpha}$ and $y_{i\alpha}$ have been drawn.

The functions L_i and U_i are called the tolerance limits for the variate x_i. We will say that L_i is the lower, and U_i the upper tolerance limit of x_i. In general, there exist infinitely many tolerance limits L_i and U_i which are solutions of Problem 1 or Problem 2. It is clear that the tolerance limits L_i and U_i are the more favorable the smaller the difference $U_i - L_i$. Hence if there exist several solutions for the tolerance limits L_i and U_i we should select that one for which the difference $U_i - L_i$ becomes a minimum in some sense.

S. S. Wilks[1] gave a solution of Problems 1 and 2 in the univariate case, i.e.

[1] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Annals of Math. Stat.*, Vol. 12 (1941). See also his paper on the same subject presented at the meeting of the Institute of Mathematical Statistics in Poughkeepsie, September, 1942.

if $p = 1$. It seems that Wilks' solution is the best possible one if nothing is known about the probability density function except that it is continuous. However, if it is known a priori that the unknown density function is an element of a k-parameter family of functions, it will in general be possible to derive tolerance limits which are considerably better than those proposed by Wilks.

Wilks' results can easily be extended to the multivariate case provided the variates $x_1, \cdots, x_p$ are known to be independently distributed[2] This is a serious restriction, since in many practical cases the independence of the variates $x_1, \cdots, x_p$ cannot be assumed. The case of dependent variates has not been treated by Wilks.

In this paper we give a solution of problems 1 and 2 when the size n of the sample is large. In the next section a lemma is proved which will be used in the derivation of tolerance limits. In section 3 the univariate case is treated and in section 4 the results are extended to the multivariate case.

2. A lemma. We will prove the following

LEMMA. *Let $\{x_{1n}\}, \cdots, \{x_{rn}\}$ ($n = 1, 2, \cdots$, ad inf) be r sequences of random variables and let $a_1, \cdots, a_r$ be r constants such that the joint distribution of $\sqrt{n}(x_{1n} - a_1), \cdots, \sqrt{n}(x_{rn} - a_r)$ converges with $n \to \infty$ towards the r-variate normal distribution with zero means and finite non-singular covariance matrix $\|\sigma_{ij}\|$ ($i, j = 1, \cdots, r$). Furthermore, let $g(u_1, \cdots, u_r)$ be a function of r variables $u_1, \cdots, u_r$ which admits continuous first derivatives in the neighborhood of the point $u_1 = a_1, \cdots, u_r = a_r$. Assume that at least one of the first partial derivatives of $g(u_1, \cdots, u_r)$ is not zero at the point $u_1 = a_1, \cdots, u_r = a_r$. Then the distribution of $\sqrt{n}[g(x_{1n}, \cdots, x_{rn}) - g(a_1, \cdots, a_r)]$ converges with $n \to \infty$ towards the normal distribution with zero mean and variance $\sigma_g^2 = \sum_j \sum_i \sigma_{ij} g_i g_j$ where g_i denotes the partial derivative of $g(u_1, \cdots, u_r)$ with respect to u_i taken at $u_1 = a_1, \cdots, u_r = a_r$.*

Proof: Since the joint distribution of $\sqrt{n}(x_{1n} - a_1), \cdots, \sqrt{n}(x_{rn} - a_r)$ approaches an r-variate normal distribution with zero means and finite non-singular covariance matrix, the probability that

$$a_i - \frac{1}{\sqrt[3]{n}} \le x_{in} \le a_i + \frac{1}{\sqrt[3]{n}} \qquad (i = 1, \cdots, r) \tag{3}$$

holds, converges to 1 with $n \to \infty$. From (3) and the continuity of the first derivatives of $g(u_1, \cdots, u_r)$ it follows easily that for any positive ϵ the probability that

$$\begin{aligned}(4)\qquad &\sum_{i=1}^{r} \sqrt{n}\,(x_{in} - a_i)g_i - \epsilon \\ &\le \sqrt{n}[g(x_{1n}, \cdots, x_{rn}) - g(a_1, \cdots, a_r)] \le \sum_i \sqrt{n}\,(x_{in} - a_i)g_i + \epsilon\end{aligned}$$

[2] This was mentioned by Wilks in his paper presented at the meeting of the Institute of Mathematical Statistics in Poughkeepsie, N. Y., September, 1942.

holds, converges to 1 with $n \to \infty$. Since the limit distribution of $\sum_i \sqrt{n}(x_{in} - a_i)g_i$ is normal with zero mean and variance equal to $\Sigma\Sigma\sigma_{ij}g_ig_j$, our Lemma follows easily from the fact that the quantity ϵ in (4) can be chosen arbitrarily small

3. The univariate case. In this section we assume that $p = 1$. Hence the probability density function $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ is replaced by the univariate density function $f(x, \theta_1, \cdots, \theta_k)$ In order to simplify the notations, the letter θ without any subscript will be used to denote the set of parameter values $\theta_1, \cdots, \theta_k$.

For any positive $\xi < 1$ let $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ be two functions of θ such that

$$\int_{\varphi(\theta, \xi)}^{\psi(\theta, \epsilon)} f(x, \theta)\, dx = \xi. \tag{5}$$

If $f(x, \theta)$ is a continuous function of x, functions $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ satisfying (5) exist. It is clear that for any function $\varphi(\theta, \xi)$ subject to the condition

$$\int_{-\infty}^{\varphi(\theta, \epsilon)} f(x, \theta)\, dx < 1 - \xi$$

there exists a function $\psi(\theta, \xi)$ such that (5) holds. We will choose $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ so that (5) is satisfied and

$$\psi(\theta, \xi) - \varphi(\theta, \xi) \leq \bar{\psi}(\theta, \xi) - \bar{\varphi}(\theta, \xi) \tag{6}$$

for any value of θ and for any functions $\bar{\varphi}(\theta, \xi)$ and $\bar{\psi}(\theta, \xi)$ which satisfy (5).

Let $\hat{\theta}_i$ $(i = 1, \cdots, k)$ be the maximum likelihood estimate of θ_i calculated from the observations $x_{11}, \cdots, x_{pn}$. We propose the use of the tolerance limits

$$L = \varphi(\hat{\theta}, \xi) \quad \text{and} \quad U = \psi(\hat{\theta}, \xi) \tag{7}$$

where the value of the constant ξ has to be properly determined. Problem 1 is solved if we can determine ξ as a function of β and γ such that

$$P\left\{\int_{\varphi(\hat{\theta}, \xi)}^{\psi(\hat{\theta}, \xi)} f(x, \theta)\, dx \geq \gamma \,\middle|\, \theta\right\} = \beta. \tag{8}$$

Problem 2 is solved if we determine ξ as a function of β, λ and N such that

$$P(M \geq \lambda N \mid \theta) = \beta \tag{9}$$

where M denotes the number of observation in the second sample which lie between the tolerance limits $\varphi(\hat{\theta}, \xi)$ and $\psi(\hat{\theta}, \xi)$. The use of tolerance limits of the form (7) seems to be well justified by the fact that the functions $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ satisfy (5) and (6) and that $\hat{\theta}_i$ is an optimum estimate of θ_i $(i = 1, \cdots, k)$.

Now we will derive the large sample distribution of

(10) $$I(\hat{\theta}, \theta, \xi) = \int_{\varphi(\hat{\theta}, \xi)}^{\psi(\hat{\theta}, \xi)} f(x, \theta)\, dx.$$

We obviously have

(11) $$I(\theta, \theta, \xi) = \xi.$$

We will assume that the limit joint distribution of $\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_k - \theta_k)$ is normal with mean values 0 and non-singular covariance matrix $\| \sigma_{ij}(\theta) \| = \| c_{ij}(\theta) \|^{-1}$ where $c_{ij}(\theta)$ denotes the expected value of $-\frac{\partial^2 \log f(x, \theta)}{\partial\theta_i \, \partial\theta_j}$ $(i, j = 1, \cdots, k)$. This is known to be true if $f(x, \theta)$ satisfies some regularity conditions.[3] Furthermore we assume that $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ admit continuous first partial derivatives with respect to $\theta_1, \cdots, \theta_k$ and that $f(x, \theta)$ is a continuous function of x in the neighborhood of $x = \varphi(\theta, \xi)$ and $x = \psi(\theta, \xi)$. We have

(12) $$\left.\frac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta} = \frac{\partial \psi(\theta, \xi)}{\partial \theta_i} f[\psi(\theta, \xi), \theta] - \frac{\partial \varphi(\theta, \xi)}{\partial \theta_i} f[\varphi(\theta, \xi), \theta]$$

Assuming that at least one of the derivatives $\left.\frac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta}$ is not zero, it follows from our Lemma that
$\sqrt{n}[I(\hat{\theta}, \theta, \xi) - I(\theta, \theta, \xi)] = \sqrt{n}[I(\hat{\theta}, \theta, \xi) - \xi]$ is in the limit normally distributed with zero mean and variance

$$\begin{aligned}
\sigma^2(\theta, \xi) = {} & \{f[\psi(\theta, \xi), \theta]\}^2 \sum_j \sum_i \frac{\partial \psi(\theta, \xi)}{\partial \theta_i} \frac{\partial \psi(\theta, \xi)}{\partial \theta_j} \sigma_{ij}(\theta) \\
& - 2f[\psi(\theta, \xi), \theta] f[\varphi(\theta, \xi), \theta] \sum_j \sum_i \frac{\partial \psi(\theta, \xi)}{\partial \theta_i} \frac{\partial \varphi(\theta, \xi)}{\partial \theta_j} \sigma_{ij}(\theta) \qquad (13) \\
& + \{f[\varphi(\theta, \xi), \theta]\}^2 \sum_j \sum_i \frac{\partial \varphi(\theta, \xi)}{\partial \theta_i} \frac{\partial \varphi(\theta, \xi)}{\partial \theta_j} \sigma_{ij}(\theta).
\end{aligned}$$

For any positive $\beta < 1$ denote by λ_β the value for which

(14) $$\frac{1}{\sqrt{2\pi}} \int_{\lambda_\beta}^{\infty} e^{-\frac{1}{2}t^2}\, dt = \beta.$$

Then the probability that

(15) $$I(\hat{\theta}, \theta, \xi) \geq \xi + \lambda_\beta \frac{\sigma(\theta, \xi)}{\sqrt{n}},$$

converges with $n \to \infty$ towards β.

Let

(16) $$\xi(\beta, \gamma, \hat{\theta}) = \gamma - \lambda_\beta \frac{\sigma(\hat{\theta}, \gamma)}{\sqrt{n}}$$

---

[3] See for instance J. L. Doob, "Probability and statistics," *Trans. Amer. Math. Soc.*, October, 1934.

If $\sigma(\theta, \xi)$ is continuous in θ and ξ, it follows easily from (15) that the probability that

$$I[\hat{\theta}, \theta, \bar{\xi}(\beta, \gamma, \hat{\theta})] \geq \gamma \tag{17}$$

holds, converges to β with $n \to \infty$. Hence we can summarize our results in the following

THEOREM 1: *Let $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ be two functions satisfying* (5) *and* (6). *Furthermore, let the functions $I(\hat{\theta}, \theta, \xi)$, $\sigma^2(\theta, \xi)$ and $\bar{\xi}(\beta, \gamma, \hat{\theta})$ be defined by* (10), (13) *and* (16) *respectively. Denote by $\theta_1^0, \cdots, \theta_k^0$ the true values of the parameters. It is assumed that there exist two positive numbers ϵ and δ such that the following three conditions are fulfilled.*

(a) *For any point θ for which $\sum_{i=1}^{k} (\theta_i - \theta_i^0)^2 \leq \epsilon$ the limit joint distribution of $\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_k - \theta_k)$, calculated under the assumption that θ is the true parameter point, is normal with zero means and a finite non-singular covariance matrix $\| \sigma_{ij}(\theta) \|$ where $\sigma_{ij}(\theta)$ is a continuous function of θ in the domain $\sum_i (\theta_i - \theta_i^0)^2 \leq \epsilon$.*

(b) *The partial derivatives $\left.\dfrac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta}$ $(i = 1, \cdots, k)$ are continuous functions of θ and ξ in the domain*

$$\sum_{i=1}^{k} (\theta_i - \theta_i^0)^2 \leq \epsilon \quad \textit{and} \quad | \xi - \gamma | \leq \delta.$$

(c) *At least one of the partial derivatives $\left.\dfrac{\partial I(\hat{\theta}, \theta^0, \gamma)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta^0}$ $(i = 1, \cdots, k)$ is not equal to zero.*

Then the probability that

$$I[\hat{\theta}, \theta^0, \bar{\xi}(\beta, \gamma, \hat{\theta})] \geq \gamma,$$

holds, converges to β with $n \to \infty$.

From Theorem 1 we obtain the following

LARGE SAMPLE SOLUTION OF PROBLEM 1. *For large n we can approximate the lower and upper tolerance limits by $\varphi[\hat{\theta}, \bar{\xi}(\beta, \gamma, \hat{\theta})]$ and $\psi[\hat{\theta}, \bar{\xi}(\beta, \gamma, \hat{\theta})]$ respectively, where $\bar{\xi}(\beta, \gamma, \hat{\theta})$ is given by* (16).

Now we will deal with Problem 2. We distinguish two cases

$$\text{(a)} \qquad \lim_{n \to \infty} \frac{N}{n} = \infty.$$

It is easy to see that in this case the solution of Problem 2 is obtained from that of Problem 1 by substituting λ for γ. Hence for large n the tolerance limits can be approximated by $\varphi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$ respectively.

For these tolerance limits condition 2 is fulfilled in the limit, i.e. $\lim_{n=\infty} P(M \geq \lambda N \mid \theta_1, \cdots, \theta_k) = \beta$

(b) The integers n and N approach infinity while $\frac{N}{n}$ remains bounded.

Denote $\sqrt{n}\,[I(\hat{\theta}, \theta, \xi) - \xi]$ by u and $\sqrt{N}\left(\frac{M(\xi)}{N} - \xi\right)$ by v, where $M(\xi)$ denotes the number of observations in the second sample which fall between the limits $\varphi(\hat{\theta}, \xi)$ and $\psi(\hat{\theta}, \xi)$. For any fixed value of u the conditional expected value of $\frac{M(\xi)}{N}$ is given by $\xi + \frac{u}{\sqrt{n}}$ and the conditional variance of $\frac{M(\xi)}{N}$ is given by $\frac{1}{N}\left(\xi + \frac{u}{\sqrt{n}}\right)\left(1 - \xi - \frac{u}{\sqrt{n}}\right)$. Hence the conditional expected value of v is equal to $u\sqrt{\frac{N}{n}}$ and the conditional variance of v is equal to $\left(\xi + \frac{u}{\sqrt{n}}\right)\left(1 - \xi - \frac{u}{\sqrt{n}}\right)$. Since the limit distribution of u is normal with zero mean and standard deviation $\sigma(\theta, \xi)$ given in (13), we find that the limit bivariate distribution of u and v is given by

$$\text{(18)} \qquad \frac{1}{2\pi\sigma(\theta, \xi)\sqrt{\xi(1 - \xi)}} \exp\left[-\frac{u^2}{2\sigma^2(\theta, \xi)} - \frac{\left(v - \sqrt{\frac{N}{n}}\,u\right)^2}{2\xi(1 - \xi)}\right] du\, dv.$$

From (18) it follows that the limit distribution of v is normal with zero mean and variance

$$\text{(19)} \qquad \begin{aligned} \sigma_v^2 &= \sigma^2(\theta, \xi)\left(\frac{1}{\sigma^2(\theta, \xi)} + \frac{N}{n\xi(1 - \xi)}\right)\xi(1 - \xi) \\ &= \frac{n\xi(1 - \xi) + N\sigma^2(\theta, \xi)}{n}. \end{aligned}$$

From (19) it follows easily that the probability that

$$\text{(20)} \qquad \frac{M(\xi)}{N} \geq \xi + \frac{\lambda_\beta \sigma_v}{\sqrt{N}}$$

converges to β with $n \to \infty$. Let

$$\text{(21)} \qquad \xi^*(\beta, \lambda, \hat{\theta}) = \lambda - \frac{\lambda_\beta}{\sqrt{N}}\sqrt{\frac{n\lambda(1 - \lambda) + N\sigma^2(\hat{\theta}, \lambda)}{n}}.$$

From (20) it follows that the probability that

$$\frac{M}{N} \geq \lambda,$$

converges to β with $n \to \infty$. The letter M denotes the number of observations in the second sample which lie between the limits $\varphi[\theta, \xi^*(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \xi^*(\beta, \lambda, \hat{\theta})]$.

We can summarize our results in the following

THEOREM 2. *Let $\varphi(\theta, \xi)$ and $\psi(\theta, \xi)$ be two functions satisfying* (5) *and* (6). *Two samples of size n and N respectively are drawn and the maximum likelihood estimate $\hat{\theta}$ is calculated from the first sample only. Assume that conditions* (a), (b) *and* (c) *of Theorem* 1 *are satisfied. Let $\bar{\xi}(\beta, \gamma, \hat{\theta})$ and $\xi^*(\beta, \lambda, \hat{\theta})$ be defined by* (16) *and* (21) *respectively.*

If n and $\frac{N}{n}$ both approach infinity, the probability that $\frac{M}{N} \geq \lambda$ holds, converges to β, where M denotes the number of observations in the second sample which lie between the limits $\varphi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$.

If n and N approach infinity while $\frac{N}{n}$ remains bounded, the probability that $\frac{M}{N} \geq \lambda$ holds, converges to β, where M denotes the number of observations in the second sample which lie between the limits $\varphi[\hat{\theta}, \xi^*(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \xi^*(\beta, \lambda, \hat{\theta})]$.

From Theorem 2 we obtain the following

LARGE SAMPLE SOLUTION OF PROBLEM 2. *If n and $\frac{N}{n}$ both approach infinity the lower and upper tolerance limits can be approximated by $\varphi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \bar{\xi}(\beta, \lambda, \hat{\theta})]$ respectively. If n and N both approach infinity while $\frac{N}{n}$ remains bounded, the tolerance limits can be approximated by $\varphi[\hat{\theta}, \xi^*(\beta, \lambda, \hat{\theta})]$ and $\psi[\hat{\theta}, \xi^*(\beta, \lambda, \hat{\theta})]$ respectively. The expressions $\bar{\xi}(\beta, \lambda, \hat{\theta})$ and $\xi^*(\beta, \lambda, \hat{\theta})$ are given by* (16) *and* (21) *respectively.*

4. The multivariate case. For any positive $\xi < 1$ let $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ $(i = 1, \cdots, p)$ be p pairs of functions of θ such that

$$\int_{\varphi_p(\theta,\xi)}^{\psi_p(\theta,\xi)} \cdots \int_{\varphi_1(\theta,\xi)}^{\psi_1(\theta,\xi)} f(x_1, \cdots, x_p, \theta)\, dx_1 \cdots dx_p = \xi. \tag{22}$$

If $f(x_1, \cdots, x_p, \theta)$ is a continuous function of $x_1, \cdots, x_p$, functions $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ $(i = 1, \cdots, p)$ satisfying (22) certainly exist. As in the univariate case, there will be infinitely many sets of p pairs of functions $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ which satisfy (22). Since we wish to have tolerance limits as narrow as possible, we will try to choose the functions $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ so that $\psi_i(\theta, \xi) - \varphi_i(\theta, \xi)$ should be as small as possible. Since it is impossible to minimize all p differences $\psi_1(\theta, \xi) - \varphi_1(\theta, \xi), \cdots, \psi_p(\theta, \xi) - \varphi_p(\theta, \xi)$ simultaneously, we will have to be satisfied with some compromise solution. For example, we could minimize the product $\prod_i [\psi_i(\theta, \xi) - \varphi_i(\theta, \xi)]$ or some other function of the p differences $\psi_i(\theta, \xi) - \varphi_i(\theta, \xi)$. Another reasonable procedure would be to minimize

$\prod_i [\psi_i(\theta, \xi) - \varphi_i(\theta, \xi)]$ subject to (22) and the condition that for any i and j, $\dfrac{\psi_i(\theta, \xi) - \varphi_i(\theta, \xi)}{\psi_j(\theta, \xi) - \varphi_j(\theta, \xi)}$ is equal to the ratio of the standard deviation of x_i to that of x_j.

Here we will deal with the problem of deriving tolerance limits for the variates $x_1, \cdots, x_p$ after the functions $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ have been chosen. Since the theory of the multivariate case is very similar to that of the univariate case, we will merely outline it briefly.

As tolerance limits for x_i we will use the functions $\varphi_i(\hat{\theta}, \xi)$ and $\psi_i(\hat{\theta}, \xi)$ where the value of ξ has to be properly determined. Problem 1 is solved if we can determine ξ as a function of β and γ so that

$$P\left\{\int_{\varphi_p(\hat{\theta},\xi)}^{\psi_p(\hat{\theta},\xi)} \cdots \int_{\varphi_1(\hat{\theta},\xi)}^{\psi_1(\hat{\theta},\xi)} f(x_1, \cdots, x_p, \theta)\, dx_1 \cdots dx_p \geq \gamma \mid \theta\right\} = \beta. \tag{23}$$

Problem 2 is solved if we determine ξ as a function of β, λ and N such that condition 2 is fulfilled. Let

$$I(\hat{\theta}, \theta, \xi) = \int_{\varphi_p(\hat{\theta},\xi)}^{\psi_p(\hat{\theta},\xi)} \cdots \int_{\varphi_1(\hat{\theta},\xi)}^{\psi_1(\hat{\theta},\xi)} f(x_1, \cdots, x_p, \theta)\, dx_1 \cdots dx_p \tag{24}$$

and let

$$\begin{aligned} I_i(\hat{\theta}, \theta, \xi, x_i) = \int_{\varphi_p(\hat{\theta},\xi)}^{\psi_p(\hat{\theta},\xi)} \cdots \int_{\varphi_{i+1}(\hat{\theta},\xi)}^{\psi_{i+1}(\hat{\theta},\xi)} \int_{\varphi_{i-1}(\hat{\theta},\xi)}^{\psi_{i-1}(\hat{\theta},\xi)} \\ \cdots \int_{\varphi_1(\hat{\theta},\xi)}^{\psi_1(\hat{\theta},\xi)} f(x_1, \cdots, x_p, \theta)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_p . \end{aligned} \tag{25}$$

We have

$$\begin{aligned} \left.\frac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta} = \sum_{s=1}^{p} \frac{\partial \psi_s(\theta, \xi)}{\partial \theta_i} I_s[\theta, \theta, \xi, \psi_s(\theta, \xi)] \\ - \sum_{s=1}^{p} \frac{\partial \varphi_s(\theta, \xi)}{\partial \theta_i} I_s[\theta, \theta, \xi, \varphi_s(\theta, \xi)]. \end{aligned} \tag{26}$$

Assuming that the partial derivatives $\left.\dfrac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta}$ $(i = 1, \cdots, k)$ are continuous functions and that $\left.\dfrac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i}\right|_{\hat{\theta}=\theta}$ is not zero for at least one value of i, it follows from our Lemma that $\sqrt{n}\,[I(\hat{\theta}, \theta, \xi) - I(\theta, \theta, \xi)] = \sqrt{n}\,[I(\hat{\theta}, \theta, \xi) - \xi]$ is in the limit normally distributed with mean value zero and variance

$$\begin{aligned} \bar{\sigma}^2(\theta, \xi) = \sum_{q=1}^{p} \sum_{s=1}^{p} \sum_{j=1}^{k} \sum_{i=1}^{k} \frac{\partial \psi_s(\theta, \xi)}{\partial \theta_i} \frac{\partial \psi_q(\theta, \xi)}{\partial \theta_j} I_s[\theta, \theta, \xi, \psi_s(\theta, \xi)] I_q[\theta, \theta, \xi, \psi_s(\theta, \xi)] \sigma_{ij}(\theta) \\ - 2 \sum_s \sum_q \sum_i \sum_j \frac{\partial \psi_s(\theta, \xi)}{\partial \theta_i} \frac{\partial \varphi_q(\theta, \xi)}{\partial \theta_j} \\ \cdot I_s[\theta, \theta, \xi, \psi_s(\theta, \xi)] I_q[\theta, \theta, \xi, \varphi_q(\theta, \xi)] \sigma_{ij}(\theta) \\ + \sum_s \sum_q \sum_j \sum_i \frac{\partial \varphi_s(\theta, \xi)}{\partial \theta_i} \frac{\partial \varphi_q(\theta, \xi)}{\partial \theta_j} \\ \cdot I_s[\theta, \theta, \xi, \varphi_s(\theta, \xi)] I_q[\theta, \theta, \xi, \varphi_q(\theta, \xi)] \sigma_{ij}(\theta) \end{aligned} \tag{27}$$

where $\| \sigma_{ij}(\theta) \|$ is the limit covariance matrix of $\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_k - \theta_k)$.

For any positive $\beta > 1$, let λ_β be the real value defined by the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\lambda_\beta}^{\infty} e^{-\frac{1}{2}t^2} \, dt = \beta. \tag{28}$$

Let

$$\bar{\zeta}(\beta, \gamma, \hat{\theta}) = \gamma - \lambda_\beta \frac{\bar{\sigma}(\hat{\theta}, \gamma)}{\sqrt{n}} \tag{29}$$

and

$$\zeta^*(\beta, \lambda, \hat{\theta}) = \lambda - \frac{\lambda_\beta}{\sqrt{N}} \sqrt{\frac{n\lambda(1 - \lambda) + N\bar{\sigma}^2(\hat{\theta}, \lambda)}{n}}. \tag{30}$$

We can easily prove the following two theorems:

THEOREM 3. *Let $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ $(i = 1, \cdots, p)$ be p pairs of functions which satisfy (22). Let the functions $I(\hat{\theta}, \theta, \xi)$, $\bar{\sigma}^2(\theta, \xi)$ and $\bar{\zeta}(\beta, \gamma, \hat{\theta})$ be defined by (24), (27) and (29) respectively. Denote by $\theta_1^0, \cdots, \theta_k^0$ the true values of the parameters $\theta_1, \cdots, \theta_k$. It is assumed that there exist two positive numbers ϵ and δ such that the following three conditions are fulfilled:*

(a) *For any point θ for which $\sum_{i=1}^{k} (\theta_i - \theta_i^0)^2 \leq \epsilon$ the limit joint distribution of $\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_k - \theta_k)$, calculated under the assumption that θ is the true parameter point, is normal with zero means and a finite non-singular covariance matrix $\| \sigma_{ij}(\theta) \|$ where $\sigma_{ij}(\theta)$ is a continuous function of θ in the domain $\sum_i (\theta_i - \theta_i^0)^2 \leq \epsilon$.*

(b) *The partial derivatives $\left. \frac{\partial I(\hat{\theta}, \theta, \xi)}{\partial \hat{\theta}_i} \right|_{\hat{\theta}=\theta}$ $(i = 1, \cdots, k)$ are continuous functions of θ and ξ in the domain $\sum_{i=1}^{k} (\theta_i - \theta_i^0)^2 \leq \epsilon$ and $| \xi - \gamma | \leq \delta$.*

(c) *At least one of the partial derivatives $\left. \frac{\partial I(\hat{\theta}, \theta^0, \gamma)}{\partial \hat{\theta}_i} \right|_{\hat{\theta}=\theta^0}$ $(i = 1, \cdots, k)$ is not equal to zero.*

Then the probability that

$$I[\hat{\theta}, \theta^0, \bar{\zeta}(\beta, \gamma, \hat{\theta})] \geq \gamma$$

holds, converges to β with $n \to \infty$.

THEOREM 4. *Let $\varphi_i(\theta, \xi)$ and $\psi_i(\theta, \xi)$ $(i = 1, \cdots, p)$ be p pairs of functions which satisfy (22). Two samples of size n and N respectively are drawn and the maximum likelihood estimate $\hat{\theta}$ is calculated from the first sample only. Assume that conditions (a), (b) and (c) of Theorem 3 are fulfilled and let $\bar{\zeta}(\beta, \gamma, \hat{\theta})$ and $\zeta^*(\beta, \lambda, \hat{\theta})$ be defined by (29) and (30) respectively. Denote by $y_{i\alpha}$ the outcome of the α-th observation on the i-th variate in the second sample.*

If n and $\frac{N}{n}$ both approach infinity, the probability that $M \geq \lambda N$ holds converges to β, where M denotes the number of different values of α for which

$$\varphi_i[\hat{\theta}, \bar{\zeta}(\beta, \lambda, \hat{\theta})] \leq y_{i\alpha} \leq \psi_i[\hat{\theta}, \bar{\zeta}(\beta, \lambda, \hat{\theta})] \qquad (i = 1, \cdots, p).$$

If n and N approach infinity while $\frac{N}{n}$ remains bounded, the probability that $M \geq \lambda N$ holds converges to β where M denotes the number of different values of α for which

$$\varphi_i[\hat{\theta}, \zeta^*(\beta, \lambda, \hat{\theta})] \leq y_{i\alpha} \leq \psi_i[\hat{\theta}, \zeta^*(\beta, \lambda, \hat{\theta})] \qquad (i = 1, \cdots, p).$$

The proofs of Theorems 3 and 4 are omitted since they are similar to the proofs of Theorems 1 and 2

From Theorem 3 we obtain the following

LARGE SAMPLE SOLUTION OF PROBLEM 1. *For large n we can approximate the lower and upper tolerance limits for x_i by $\varphi_i[\hat{\theta}, \bar{\zeta}(\beta, \gamma, \hat{\theta})]$ and $\psi_i[\hat{\theta}, \bar{\zeta}(\beta, \gamma, \hat{\theta})]$ respectively where $\bar{\zeta}(\beta, \gamma, \theta)$ is given by* (29).

From Theorem 4 we obtain the following

LARGE SAMPLE SOLUTION OF PROBLEM 2. *If n and $\frac{N}{n}$ approach infinity, the lower and upper tolerance limits for x_i can be approximated by $\varphi_i[\hat{\theta}, \bar{\zeta}(\beta, \lambda, \hat{\theta})]$ and $\psi_i[\hat{\theta}, \bar{\zeta}(\beta, \lambda, \hat{\theta})]$ respectively. If n and N both approach infinity while $\frac{N}{n}$ remains bounded, the tolerance limits for x_i can be approximated by $\varphi_i[\hat{\theta}, \zeta^*(\beta, \lambda, \hat{\theta})]$ and $\psi_i[\hat{\theta}, \zeta^*(\beta, \lambda, \hat{\theta})]$ respectively. The expressions $\zeta(\beta, \lambda, \hat{\theta})$ and $\zeta^*(\beta, \lambda, \hat{\theta})$ are defined in* (29) *and* (30) *respectively.*

5. An example. Let x be a normally distributed variate with mean value θ_1 and standard deviation θ_2, i.e. the probability density function of x is given by

$$f(x, \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\,\theta_2} e^{-\frac{1}{2}(x-\theta_1)^2/\theta_2^2}.$$

For any positive $\xi < 1$ let $\rho(\xi)$ be the value for which

$$\frac{1}{\sqrt{2\pi}} \int_{-\rho(\xi)}^{\rho(\xi)} e^{-\frac{1}{2}t^2}\, dt = \xi.$$

Then the functions

$$\varphi(\theta, \xi) = \theta_1 - \rho(\xi)\theta_2$$

and

$$\psi(\theta, \xi) = \theta_1 + \rho(\xi)\theta_2$$

satisfy conditions (5) and (6).

We have

$$\hat{\theta}_1 = \frac{x_1 + \cdots + x_n}{n} = \bar{x} \quad \text{and} \quad \hat{\theta}_2 = \sqrt{\frac{\sum_{\alpha=1}^{n} (x_\alpha - \bar{x})^2}{n}}.$$

The variance of $\sqrt{n}(\hat{\theta}_1 - \theta_1)$ is equal to θ_2^2 and the limit variance of $\sqrt{n}(\hat{\theta}_2 - \theta_2)$ is equal to $\frac{1}{2}\theta_2^2$. Since the covariance of $\hat{\theta}_1$ and $\hat{\theta}_2$ is equal to zero, we obtain from (13)

$$\sigma^2(\theta, \xi) = 2\left\{\frac{1}{\sqrt{2\pi}\,\theta_2} e^{-\frac{1}{2}[\rho(\xi)]^2}\right\}\{\theta_2^2 + \tfrac{1}{2}\theta_2^2[\rho(\xi)]^2\}$$

$$- 2\left\{\frac{1}{\sqrt{2\pi}\,\theta_2} e^{-\frac{1}{2}[\rho(\xi)]^2}\right\}\{\theta_2^2 - \tfrac{1}{2}\theta_2^2[\rho(\xi)]^2\}$$

$$= \frac{1}{\pi}[\rho(\xi)]^2 e^{-[\rho(\xi)]^2}.$$

Hence for large n the tolerance limits satisfying (1) can be approximated by $\hat{\theta}_1 - \rho(\bar{\xi})\hat{\theta}_2$ and $\hat{\theta}_1 + \rho(\bar{\xi})\hat{\theta}_2$ respectively where

$$\bar{\xi} = \gamma - \lambda_\beta \frac{\rho(\gamma)}{\sqrt{n\pi}} e^{-\frac{1}{2}[\rho(\gamma)]^2}$$

and λ_β is the value determined by the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\lambda_\beta}^{\infty} e^{-\frac{1}{2}t^2}\,dt = \beta.$$

If n and N are large, the tolerance limits satisfying (2) can be approximated by $\hat{\theta}_1 - \rho(\xi^*)\hat{\theta}_2$ and $\hat{\theta}_1 + \rho(\xi^*)\hat{\theta}_2$ respectively where

$$\xi^* = \lambda - \lambda_\beta \sqrt{\frac{\lambda(1-\lambda)}{N} + \frac{[\rho(\lambda)]^2}{n\pi} e^{-[\rho(\lambda)]^2}}.$$

STATISTICAL PREDICTION WITH SPECIAL REFERENCE TO THE PROBLEM OF TOLERANCE LIMITS[1]

By S. S. Wilks

Princeton University

1. Introduction. Statistical methodology is becoming recognized in industry as an effective tool for dealing with certain problems of inspection and quality control in mass production. Quality control experts have found statistical methods useful in detecting excessive variation in a given quality characteristic of a product from a series of observations on the given quality characteristic, and in isolating the causes of such variations back in the materials or operations involved in manufacturing the product. By a process of successive detection and elimination of causes of variability, a *controlled state of quality* is established. A practical statistical procedure for establishing a controlled state of quality has been developed by Shewhart.[2] More recently, manuals for routine application of this procedure have been issued by the American Standards Association.[3]

In this paper we do not propose to go into a discussion of the application of the well known Shewhart procedure. The reader may refer to the literature mentioned in footnotes 2 and 3 for such discussion. It is sufficient to remark that experience shows that the application of this procedure leads to a controlled state of quality. Such a state of control provides a basis for making statistical predictions about measurements on the given quality characteristic in future production.

More specifically, suppose a given quality characteristic of a given product is measured by a variable X, such that X has a specific value for each individual product-piece. For example, the product may be a given type of fuse and X may be the blowing time in seconds. A product-piece would be a single fuse, and X would take on a value for each fuse. Thus, for a sequence of n fuses taken from the production line, there would be a corresponding sequence of values of X, say $X_1, X_2, \cdots X_n$. If a state of control has been established with respect to blowing time as measured by X, then the sequence of values of X will "behave like a random sequence." By this we mean that the sequence will be such that we can safely assume that it can be described mathematically by regarding X as a *continuous random variable*, i.e., such that there exists some

---

[1] An expository paper presented at a joint session of the American Mathematical Society and the Institute of Mathematical Statistics at Poughkeepsie, September 9 1942.

[2] W. A. Shewhart, *Control of Quality of Manufactured Product*, D. Van Nostrand Company, New York, 1931.

[3] *Guide for Quality Control and Control Chart Method of Analyzing Data (1941)*, and *Control Chart Method of Controlling Quality During Production (1942)*, American Standards Association, New York.

probability function $f(x)$ which describes the distribution of values of X, such that $\int_a^b f(x)\,dx$ is the probability that $a < X < b$ for any two real numbers a and b. Now, suppose we consider a sequence or *sample* S_1 of n values of X, and let X_1 and X_n be the smallest and largest values of X in the sequence. The types of questions with which we are concerned are the following: If a further sample, say S_2 of N values of X is taken, what is the probability P that at least N_0 of the values will lie between X_1 and X_2 as determined by S_1? If we choose a given probability α, at least what proportion of values of X in an indefinitely large sample S_2 will fall between X_1 and X_2 of S_1 with probability α? What is the probability P' that at least N_0 of the values of S_2 will exceed X_1 of S_1? At least what proportion of values of X in an indefinitely large sample S_2 will exceed X_1 with probability α? These questions suggest several of a more general nature which can be treated by methods similar to those which will be discussed. For example, instead of taking X_1 and X_n, i.e. the smallest and largest items in S_1 as *tolerance limits* we could use X_m and X_{n-m+1}. More generally, we may define $100R_\alpha\%$ *tolerance limits* $L_1(x_1, x_2, \cdots x_n)$ *and* $L_2(x_1, x_2, \cdots, x_n)$ *for probability level* α of a sample S_1 of size n from a population with distribution $f(x)\,dx$ as two functions of the X's in S_1 such that the probability is α that at least $100R_\alpha\%$ of the X's of a further indefinitely large sample S_2 (i.e. the population) will lie between L_1 and L_2. Or more briefly

$$P\left(\int_{L_1}^{L_2} f(x)\,dx \geq R_\alpha\right) = \alpha.$$

The same notion clearly applies if S_2 is a finite sample of size N, rather than an indefinitely large one. In this case we would be interested in the largest integer N_α such that the probability is at least α that at least $100\bar{R}_\alpha\%$ $\left(\bar{R}_\alpha = \frac{N_\alpha}{N}\right)$ of the X's in S_2 would lie between L_1 and L_2. In most practical situations we are able to assume nothing more about $f(x)$ than it is a probability density function. We make only this assumption here. The only functions of the values of X in S_1 that we shall consider here in setting tolerance limits are *order statistics*, i.e. the ordered values of X, because the results will then be fairly simple and independent of $f(x)$.

2. A General Probability Formula. It will be convenient perhaps to derive a general probability formula at this stage from which we can derive certain special cases as we need them.

Let $X_1, X_2, \cdots, X_n$ be the n values of X in S_1 arranged in order of increasing magnitude. Let $r_1, r_2, \cdots, r_k$ be integers such that $1 \leq r_1 < r_2 < \cdots < r_k \leq n$. Let $x_{r_1}, x_{r_2}, \cdots, x_{r_k}$ be k real numbers. Let

$$\int_{-\infty}^{x_{r_1}} f(x)\,dx = p_1, \int_{x_{r_1}}^{x_{r_2}} f(x)\,dx = p_2, \cdots, \int_{x_{r_k}}^{\infty} f(x)\,dx = p_{k+1},$$

from which

$$f(x_{r_1})\,dx_{r_1} = dp_1\,, \qquad f(x_{r_2})\,dx_{r_2} = dp_2\,, \cdots, f(x_{r_k})\,dx_{r_k} = dp_k\,.$$

Then assuming $X_1, X_2, \cdots, X_n$ to be a random sample (ordered) from a population with probability element $f(x)\,dx$ it follows from the multinomial distribution law[4] that the probability of $x_{r_i} < X_{r_i} < x_{r_i} + dx_{r_i}$ $(i = 1, 2, \cdots, k)$ is given by

$$(1) \qquad \frac{n!}{r_1 - 1!\,r_2 - r_1 - 1!\cdots r_k - r_{k-1} - 1!\,n - r_k!}\, p_1^{r_1-1} p_2^{r_2-r_1-1} \cdots p_k^{r_k-r_{k-1}-1} p_{k+1}^{n-r_k}\, dp_1 dp_2 \cdots dp_k$$

except for terms of order higher than $(dp_1 dp_2 \cdots dp_k)$. Given that $X_{r_1} = x_{r_1}, \cdots, X_{r_k} = x_{r_k}$ in S_1, the conditional probability that $N_1, N_2, \cdots, N_{k+1}$ $\left(\sum_1^{k+1} N_i = N\right)$ of the values of X in S_2 will fall in the intervals $(-\infty, x_{r_1})$, $(x_{r_1}, x_{r_2}), \cdots, (x_{r_k}, \infty)$ respectively is by the multinomial law

$$(2) \qquad \frac{N!}{N_1!\,N_2!\cdots N_{k+1}!}\, p_1^{N_1} p_2^{N_2} \cdots p_{k+1}^{N_{k+1}}\,.$$

The joint probability law of $X_{r_1}, X_{r_2}, \cdots X_{r_k}$ and $N_1, N_2, \cdots, N_{k+1}$ $\left(\sum_1^{k+1} N_i = N\right)$ is given by the product of (1) and (2). Integrating this product with respect to the x's (i.e. the p's) we find the probability law of the N's to be

$$(3) \qquad \frac{N!\,n!\,N_1 + r_1 - 1!\,N_2 + r_2 - r_1 - 1!\cdots N_k + r_k - r_{k-1} - 1!\,N_{k+1} + n - r_k!}{r_1 - 1!\,r_2 - r_1 - 1!\cdots r_k - r_{k-1} - 1!\,n - r_k!\,N + n!\,N_1!\,N_2!\cdots N_{k+1}!}$$

which is clearly independent of $f(x)$. This result can be derived by direct combinatorial methods but the present derivation provides a simple proof that the result is independent of $f(x)$.

3. The Problem of One Tolerance Limit. There are problems in quality control in which it is important to consider only one tolerance limit. For example, in testing breaking strength of steel wire the most significant tolerance limit is the lower one. The problem of prediction in this case is as follows:

[4] Which states that if a trial results in one and only one of the mutually exclusive events $E_1, E_2, \cdots, E_k$, the probability P that in a total of n trials n_1 will result in E_1, n_2 in $E_2, \cdots, n_k$ in E_k $\left(\sum_1^k n_i = n\right)$, is given by

$$P = \frac{n!}{n_1!\,n_2!\cdots n_k!}\, p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

where $p_1, p_2, \cdots, p_k$, $\left(\sum_1^k p_i = 1\right)$ are the probabilities of a single trial resulting in E_1, $E_2, \cdots, E_k$ respectively.

Suppose the given quality characteristic, as measured by X, is in a state of statistical control, and that a sequence of n measurements on X have been made. Let X_1 be the smallest of the n values. What is the probability that at least N_0 of N further measurements on X will exceed the value X_1 as determined by the initial sample? Instead of considering the smallest value of X as the lower tolerance limit we could just as easily choose the second smallest, or any other small *order statistic* but the case of the smallest value is perhaps of greater practical interest than any other case. The problem of an upper tolerance limit is entirely similar to that of a lower tolerance limit.

TABLE I

Values of N_α and $\bar{R}_\alpha$ for $\alpha = 0.99$ and 0.95 for several combinations of values of N and n, and for the problem of one tolerance limit. (For $N = \infty$, $\bar{R}_\alpha$ is denoted by R_α)

n	N	$\alpha = 0.99$		$\alpha = 0.95$	
		N_{99}	$\bar{R}_{99}$	N_{95}	$\bar{R}_{95}$
10	10	5	.500	7	.700
10	20	11	.550	14	.700
10	∞	—	.631	—	.741
50	50	44	.880	46	.920
50	100	90	.900	93	.930
50	∞	—	.912	—	.942
100	100	94	.940	96	.960
100	200	189	.945	193	.965
100	∞	—	.955	—	.970
500	500	494	.988	496	.992
500	1000	989	.989	993	.993
500	∞	—	.991	—	.994

The probability $P_1(N_0)$ that N_0 of the N further measurements will exceed the smallest value of X in an initially drawn sample of size n is given by (3) for $k = 1$, $r_1 = 1$, $N_2 = N_0$, $N_1 = N - N_0$, i.e.

$$P_1(N_0) = n \frac{N!\,(N_0 + n - 1)!}{N_0!\,(N + n)!}. \tag{4}$$

Values of $P_1(N_0)$ can be easily calculated by using the recursion formula

$$P_1(N_0 - 1) = \frac{N_0}{N_0 + n - 1} P_1(N_0). \tag{5}$$

For given values of N, n and α we are interested in the largest integer N_α for which

$$\sum_{N_0=N_\alpha}^{N} P_1(N_0) \geq \alpha. \tag{6}$$

If we set $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and set $\lim_{N\to\infty} \bar{R}_\alpha = R_\alpha$ it can be verified that the value of R_α is given by solving the following equation for R_α

$$n \int_{R_\alpha}^{1} \xi^{n-1}\, d\xi = \alpha. \tag{7}$$

It will be observed that $n\xi^{n-1}\, d\xi$ is to within terms of order $d\xi$ the probability that $\xi < \int_{X_1}^{\infty} f(x)\, dx < \xi + d\xi$ in samples of size n from a distribution with probability element $f(x)\, dx$, where X_1 is the smallest value of X in the sample. The statistical interpretation of (7) is simply this: *The probability is α that the proportion of values of X exceeding X_1 in a further indefinitely large sample is at least R_α.*

Choosing $\alpha = 0.99$ and 0.95 Table I shows values of N_α and $\bar{R}_\alpha$ for various combinations of values of n and N for the case of one tolerance limit. The table indicates the degree of precision with which predictions about a single tolerance limit can be made from a sample of size n about a further sample of size N for a few important values of n and N. It should be noted that each prediction is made concerning a pair of samples, i.e. an initial sample of size n and a further sample of size N and that the prediction holds for any function $f(x)$. Thus as a typical entry we may state that if a sample of 100 is drawn and also a sample of 200, then the probability is 0.99 (approx.) that the X's of at least 189 (or 94.5%) of the cases in the second sample will exceed the smallest X in the first sample.

4. The Problem of Two Tolerance Limits. Again, suppose the given quality characteristic as measured by X is in a state of statistical control and that a sequence of n measurements are made on X. Let X_1 and X_n be the smallest and largest values of X respectively. The question to be considered now is the following: What is the probability that at least N_0 of N further measurements on X will lie between the values X_1 and X_n, as determined by the initial sample?

We proceed by considering the special case of (3) for which $k = 2$, $r_1 = 1$ $r_2 = n$, $N_2 = N_0$, $N_3 = N - N_0 - N_1$. We find for the joint distribution of N_1 and N_0

$$P(N_1, N_0) = \frac{N!\, n!\, (N_0 + n - 2)!}{(n-2)!\, N_0!\, (N+n)!}. \tag{8}$$

To obtain the distribution of N_0, we simply sum (8) with respect to N_1 from 0 to $N - N_0$, thus obtaining

$$(9) \qquad P_2(N_0) = n(n-1)(N - N_0 + 1)\,\frac{N!\,\overline{N_0 + n - 2}!}{N_0!\,\overline{N+n}!}.$$

A convenient recursion formula for computation purposes is

$$(10) \qquad P_2(N_0 - 1) = \frac{N_0(N - N_0 + 2)}{(N - N_0 + 1)(N_0 + n - 2)}\,P_2(N_0).$$

For given values of N, n and α we require the largest value of N_α for which

$$(11) \qquad \sum_{N_0 = N_\alpha}^{N} P_2(N_0) \geq \alpha.$$

Setting $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and $\lim_{N \to \infty} \bar{R}_\alpha = R_\alpha$ one finds that R_α is given by solving the equation[5] for R_α

$$(12) \qquad n(n-1)\int_{R_\alpha}^{1} \xi^{n-2}(1 - \xi)\,d\xi = \alpha.$$

It can be verified that $n(n-1)\xi^{n-2}(1-\xi)\,d\xi$ is to within terms of order $d\xi$ the probability that $\xi < \int_{X_1}^{X_n} f(x)\,dx < \xi + d\xi$, thus showing that (12) is the probability that the proportion of an indefinitely large number of further values of X lying between X_1 and X_n is at least R_α.

Table II gives, for the case of two tolerance limits, values of N_α and R_α for several important combinations of n and N, including limiting values R_α of $\bar{R}_\alpha$ for indefinitely large N.

It should be noted that the problem of two tolerance limits can be immediately extended to the case where the lower and upper tolerance limits may be any two of the order statistics in S_1.

5. The Problem of Tolerance Limits for Two Quality Characteristics. We have thus far devoted our discussion to the problem of tolerance limits for a single quality characteristic. The problem of two or more quality characteristics can be treated by methods similar to those already used. The simplest case is that in which each product-piece under consideration is measured on two independent quality characteristics. Suppose the two characteristics are measured by X and Y. Let a sample of n product-pieces be taken, assuming a state of statistical control has been established, and let X_1 be the smallest of the X values and Y_1 the smallest of the Y values. The question with which we are

[5] This limiting case in the problem of tolerance limits as well as that expressed in (7) and other similar limiting cases have been considered by the author in an earlier paper "Determination of Sample Sizes for Setting Tolerance Limits," *Annals of Math. Stat.* Vol. XII (1941) pp. 91–96.

concerned here is the following: If N further product-pieces are measured on X and Y, what is the probability that $X > X_1$ and $Y > Y_1$ for N_0 of the pieces? Let X and Y be statistically independent and let $f(x)$ and $g(y)$ be the probability functions of X and Y respectively. Let $\int_{-\infty}^{X_1} f(x)\,dx = p$ and $\int_{-\infty}^{Y_1} g(y)\,dy = q$. The probability law of p and q is

$$n^2(1 - p)^{n-1}(1 - q)^{n-1}\,dp\,dq. \tag{13}$$

TABLE II

Values of N_α and $\bar{R}_\alpha$ for $\alpha = .99$ and $.95$ for several combinations of values of N and n and for the problem of two tolerance limits. (For $N = \infty$, $\bar{R}_\alpha$ is denoted by R_α)

n	N	$\alpha = 0.99$		$\alpha = 0.95$	
		$N_{.99}$	$\bar{R}_{.99}$	$N_{.95}$	$\bar{R}_{.95}$
10	10	4	.400	5	.500
10	20	8	400	11	.550
10	∞	—	.496	—	.606
50	50	42	.840	44	.880
50	100	85	.850	90	.900
50	∞	—	.874	—	.909
100	100	89	.890	92	.920
100	200	184	.920	188	.940
100	∞	—	.935	—	.953
500	500	491	.982	494	.988
500	1000	985	.985	989	.989
500	∞	—	.987	—	.991

In a further sample of size N the probability that for N_0 of the cases, $X > X_1$ and $Y > Y_1$, X_1 and Y_1 being determined by the first sample, is

$$\frac{N!}{N_0!\,N - N_0!}\,[(1 - p)(1 - q)]^{N_0}[1 - (1 - p)(1 - q)]^{N-N_0}. \tag{14}$$

The joint probability law of N_0, p and q is given by the product of (13) and (14). Integrating this product with respect to p and q we obtain as the probability law of N_0,

$$H_3(N_0) = n^2 \binom{N}{N_0} \sum_{i=0}^{N-N_0} \binom{N - N_0}{i} \frac{(-1)^i}{(n + N_0 + i)^2}. \tag{15}$$

For given values of N, n and α it is important, as before, to determine N_α as the largest integer for which

$$\sum_{N_0=N_\alpha}^{N} P_3(N_0) \geq \alpha. \tag{16}$$

Setting $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and $\lim_{N\to\infty} \bar{R}_\alpha = R_\alpha$ one finds R_α to be given by solving the following equation for R_α

$$-n^2 \int_{R_\alpha}^{1} \xi^{n-1} \log \xi \, d\xi = \alpha \tag{17}$$

The expression $-n^2\xi^{n-1} \log \xi \, d\xi$ is simply the probability that $\xi < \left(\int_{X_1}^{\infty} f(x)\, dx\right)\left(\int_{Y_1}^{\infty} g(y)\, dy\right) < \xi + d\xi$ to within terms of order $d\xi$, which is the proportion of the population pairs (X, Y) for which $X > X_1$ and $Y > Y_1$.

In the problem of two tolerance limits for each quality characteristic, as determined by an initial sample of size n, we calculate the probability that N_0 members of a further sample of size N will fall within the two sets of tolerance limits, with respect to the two characteristics. The problem is similar to that for one tolerance limit for each of two quality characteristics. For this case, we find corresponding to (15), (16), (17), respectively, the following:

$$P_4(N_0) = n^2(n-1)^2 \binom{N}{N_0} \sum_{i=0}^{N-N_0} \binom{N-N_0}{i} \frac{(-1)^i}{(N_0+n-1+i)^2(N_0+n+1)^2}, \tag{18}$$

and

$$\sum_{N_0=N_\alpha}^{N} P_4(N_0) \geq \alpha \tag{19}$$

and

$$n^2(n-1)^2 \int_{R_\alpha}^{1} \xi^{n-2} [2(\xi - 1) - (\xi + 1) \log \xi] \, d\xi = \alpha. \tag{20}$$

The derivations of results analogous to (15), (16), (17), (18), (19), (20) for tolerance limits defined by other order statistics than least and greatest and also for more than two independent[6] quality characteristics are straightforward.

6. Further Remarks and Discussion. For a given set of tolerance limits on a random variable X as determined by an initial sample of size n, we have discussed the problem of predicting, with a given degree of probability, at least what proportion of values of x in a further sample (finite or indefinitely large) will lie between these tolerance limits. We have obtained theoretical results

---

[6] In a paper to appear in a forthcoming issue of the *Annals of Math Stat.*, A. Wald has shown how to set up tolerance limits for the case of two or more statistically *dependent* variables

which depend only on the assumption that X is a continuous random variable with some probability element $f(x)\,dx$, where $f(x)$ is not assumed known.

It should be emphasized that the concept of a random variable is very broad in the sense that X may be a random variable determined as a result of calculations on other random variables. For example, X may be the difference, product, or ratio of two random variables, or the average or any other "reasonable" function of several random variables which may be of interest in any given situation. Thus, on the basis of an initial sample of differences of two random variables, we may set up tolerance limits of differences and make predictions, for a given probability level as to how many differences in a further sample of differences will lie between these tolerance limits. Similarly for products, ratios, and other functions of random variables.

From the point of view of practical application, we should again note that the mathematical assumption that X is a random variable means that a state of statistical control as described in §1 must exist in the measurements to which the tolerance limit prediction theory is to be applied. In practice X is often a discrete variable, i.e. one which can take on only certain isolated values. For example, if X is the number of defective product-pieces in a drawing of one product-piece, X is either 0 or 1, depending on whether the piece was non-defective or defective. Our theory would not be applicable to such a case. However, if we take as a new variable the average value of X for several product-pieces, we then obtain a variable that is continuous enough for the tolerance limit theory to be applicable for all practical purposes.

Finally, we remark that although we have used, as concrete examples, situations in mass production engineering, the notions of tolerance limits and predictions within tolerance limits which have been discussed apply equally well to situations in any branch of applied science where measurements are made and used as a basis for predictions concerning future measurements.

7. Summary. After a state of statistical control has been established with respect to a quality characteristic of product-pieces in mass production by the standard statistical quality control methods developed and refined by Shewhart and others, there remains the problem of determining the accuracy of predictions as to how many future product-pieces will fall within tolerance limits specified by measurements on product-pieces already produced under the given state of control. This problem and some of its extensions are discussed in the present paper.

More specifically, suppose an initial sample of n product-pieces, manufactured under a given state of statistical control, are measured with respect to a given quality characteristic. Let X be a variable which measures the given characteristic, so that X has a definite value for each product-piece. Let X_1 be the smallest and X_n the largest value of X which occurs in the initial sample. Now consider a further sample of size N. The following problems of prediction relating to the second sample from information yielded by the initial sample are

considered: (1) What is the probability that at least N_0 values of X in the second sample will exceed the *tolerance limit* X_1 set by the first sample? (2) What is the probability that at least N_0 values of X in the second sample will lie between the two *tolerance limits* X_1 and X_n set by the first sample? (3) For given values of n and N and α (e.g., .99 or .95), what is the largest integer N_α such that the probability is at least α that $N_0 \geq N_\alpha$? (4) What is the limiting value of $\frac{N_\alpha}{N} = \bar{R}_\alpha$ as N increases indefinitely? Tables of values of N_α and R_α are given for each of the two problems (1) and (2), for several important combinations of values of n and N and for $\alpha =$ 99 and .95.

Problems similar to (1), (2) and (3) are discussed for the case in which tolerance limits are placed on two or more quality characteristics simultaneously.

The generality of the theory of tolerance limits and how it applies to differences, products and ratios and other functions of two or more random variables are briefly discussed.

GENERALIZED POISSON DISTRIBUTION

By F. E. Satterthwaite

Aetna Life Insurance Company

1. Introduction. The Poisson distribution is one of the most fundamental of statistical distributions. It is the distribution law for the number of events if the probability of an event happening in any infinitesimal unit of time is independent of the probability of its happening in any other unit of time. Frequently when we analyze statistics which obey the Poisson law it is desirable to give varying weights to the different events instead of considering them all of equal value. Such is the case in analyzing insurance statistics where the events are the claims received by the office and the weights are the cost of the claim to the company. We shall now show how the Poisson distribution can be generalized so as to be adequate for such an analysis.

2. First development. Let $f(x, \alpha)$ be the distribution function of the weights assigned to the events where the variable, x, refers to the weight and the variable, α, refers to time. The characteristic function of $f(x, \alpha)$ is

$$\phi(t, \alpha) = \int e^{itx} f(x, \alpha)\, dx.$$

Also let $p(\alpha)\, d\alpha$ be the probability that an event will occur in the infinitesimal unit of time, α to $\alpha + d\alpha$. If y represents the sum of the weights, the distribution function of y for this unit of time is

$$\begin{aligned} F_{d\alpha}(y, \alpha) &= 1 - p(\alpha)\, d\alpha, \qquad & y = 0 \\ &= f(y, \alpha)p(\alpha)\, d\alpha, \qquad & y > 0. \end{aligned} \tag{1}$$

The characteristic function of this distribution is

$$\begin{aligned} \Phi_{d\alpha}(t, \alpha) &= e^{it0}(1 - p(\alpha)\, d\alpha) + p(\alpha)\, d\alpha \int e^{ity} f(y, \alpha)\, dy \\ &= 1 - p(\alpha)\, d\alpha(1 - \phi(t, \alpha)) \\ &= e^{-p(\alpha)d\alpha(1-\phi(t,\alpha))}. \end{aligned} \tag{2}$$

In forming equations (1) and (2) we ignore infinitesimals of orders higher than the first in the $d\alpha$.

The expected number of events in the period of time from α_1 to α_2 is

$$P = \int_{\alpha_1}^{\alpha_2} p(\alpha)\, d\alpha,$$

and the mean distribution of weights during the same period of time is

$$f(x) = \int_{\alpha_1}^{\alpha_2} [p(\alpha)/P] f(x, \alpha)\, d\alpha.$$

The characteristic function of this mean distribution of weights is

$$\phi(t) = \int e^{itx} f(x)\, dx$$
$$= \int [p(\alpha)/P]\phi(t, \alpha)\, d\alpha.$$

These equations are based on the assumption that the probability of an event occurring in any unit of time is independent of the probability of its occurrence in any other unit of time and also the assumption that the weights assigned to each event are independent. These assumptions are implied in all that follows.

Since the characteristic function of the sum of independent variables is equal to the product of the respective characteristic functions, the characteristic function of the sum of the weights during the period of time, α_1 to α_2, is

$$\begin{aligned} \Phi(t) &= \Pi\Phi_{d\alpha}(t, \alpha) \\ &= e^{-\int p(\alpha)d\alpha + \int p(\alpha)\phi(t,\alpha)d\alpha} \qquad (3) \\ &= e^{-P(1-\phi(t))}. \end{aligned}$$

Applying the Fourier transformation, the distribution function of the sum of the weights is

$$F(y) = \frac{1}{2\pi} \int e^{-ity - P(1-\phi(t))}\, dt.$$

Equation (3) gives a convenient method for defining a generalized Poisson distribution. Any distribution which has a characteristic function in the form of $\Phi(t)$ where $\phi(t)$ is the characteristic function of an arbitrary distribution will have all the properties of a generalized Poisson distribution

3. Second development. If we let $\phi(t)$ represent the characteristic function of an arbitrary distribution, the characteristic function of the sum of n independent items obeying such a distribution law is $\Phi_n(t) = [\phi(t)]^n$. If instead of considering n to be a fixed quantity we assume that it is an independent statistical variable obeying the Poisson distribution law with mean P, the characteristic function of the sum, y, of the items of the sample becomes

$$\begin{aligned} \Phi(t) &= \Sigma_n \frac{1}{n!} P^n[\phi(t)]^n e^{-P} \\ &= e^{-P(1-\phi(t))}. \end{aligned}$$

Therefore y is seen to obey the generalized Poisson distribution law.

4. Properties. The generalized Poisson distribution preserves the unique and very important property of the Poisson distribution that nowhere in its development is it necessary to make any assumptions regarding homogeneity. The

only requirement is that the occurrence of and weight assigned to any event shall be independent of the occurrence of or weight assigned to any other event.

The distribution of the sum of the weights is a function of the expected number of events, P, and of the mean distribution of weights, $f(x)$, alone. It is independent of the way in which P and $f(x)$ are made up. Thus, if we are studying the distribution of the sum of the weights over a period of a year and if P and $f(x)$ vary with the seasons, the distribution of y is no different than it would be if P and $f(x)$ were constant. It is only necessary that the $f(x)$'s for the different seasons be weighted in proportion to the expected number of events in determining the mean $f(x)$.

Note also that in the first development it is not necessary that the variable, α, refer to time. It could just as well refer to different classes of events distinguished on any other basis. Therefore, heterogeneous material may be combined in an analysis if it is possible to determine the appropriate mean distribution of weights.

For a given weight distribution the generalized Poisson distribution for an expected number of events, nP, is identical with the distribution of the sum of n independent items each of which obeys a generalized Poisson distribution with P expected events.

Because of the property described in the preceding paragraph it is immediately apparent that a generalized Poisson distribution obeys the law of large numbers. As the number of expected events increases the distribution approaches the normal distribution.

5. Moments. The moments of a generalized Poisson distribution are functions of the moments of the underlying weight distribution. By differentiating the characteristic function we obtain the following formulas in which the presubscript, $_0$, refers to the moments of the weight distribution, $f(x)$:

$$\mu_1' = P_0\mu_1' = m$$

$$\mu_2 = P_0\mu_2' = \sigma^2$$

$$\mu_3 = P_0\mu_3'$$

$$\mu_4 = P_0\mu_4' + 3(P_0\mu_2')^2.$$

The above formulas may be verified through general reasoning by considering the moments of the distribution, $F_{d\alpha}(y, \alpha)$ (see equation (1)). This distribution refers to an infinitesimal unit of time and all the moments about zero are infinitesimals of the first order. In passing from the moments about zero to the moments about the mean the corrections are all infinitesimals of at least the second order. Therefore, the corrections may be ignored and the moments about the mean may be considered to be equal to those about zero. The above formulas follow if we take a sample of size $P/pd\alpha$ from this population.

In order to obtain Pearson's moment functions for a generalized Poisson distribution for any given mean value it is convenient to calculate the following parameters of the weight distribution:

$$
(4) \qquad \begin{aligned}
{}_0m &= {}_0\mu_1' \\
{}_0\sigma^2 &= {}_0\mu_2'/{}_0m \\
{}_0\beta_1 &= ({}_0\mu_3'/{}_0m)^2/{}_0\sigma^6 \\
{}_0(\beta_2 - 3) &= ({}_0\mu_4'/{}_0m)/{}_0\sigma^4.
\end{aligned}
$$

The Pearson moment functions then take the convenient forms:

$$
(5) \qquad \begin{aligned}
\sigma^2/m^2 &= {}_0\sigma^2/m \\
\beta_1 &= {}_0\beta_1/m \\
(\beta_2 - 3) &= {}_0(\beta_2 - 3)/m.
\end{aligned}
$$

6. Further generalizations. Often the expected number of events is not known but can be estimated to a greater or less degree of accuracy. In such a case it is convenient to assume that P is a statistical variable distributed about some expected value, say P'. A Type III distribution,

$$
g(P) = \frac{1}{\Gamma(b)}\left(\frac{b}{P'}\right)^b P^{b-1} e^{-bP/P'},
$$

will generally be as satisfactory as any to assume for P. The parameter, b, can be chosen to give any desired standard deviation. The characteristic function of the distribution of the sum of the weights under these conditions becomes

$$
\begin{aligned}
\Phi'(t) &= \int e^{-P(1-\phi(t))} g(P)\, dP \\
&= \left[1 + \frac{P'(1 - \phi(t))}{b}\right]^{-b}.
\end{aligned}
$$

The second development suggests another generalization. Instead of assuming that the number of events, n, is distributed in accord with the Poisson distribution, we may assume any discrete, non-negative distribution, $h(n)$. The distribution function for the sum of the weights is then

$$
F'(y) = \Sigma h(n) f(y, n)
$$

where $f(y, n)$ is the distribution function for the sum of n independent weights. The variance, σ^2, of this distribution is given by the formula,

$$
\frac{\sigma^2}{m^2} = \frac{{}_n\sigma^2}{{}_nm^2} + \frac{1}{{}_nm}\frac{{}_0\sigma^2}{{}_0m^2},
$$

where m refers to the mean, n refers to the distribution $h(n)$, and ${}_0$ refers to the weight distribution. Some writers have assumed that statistics of this type are distributed as a product. Such an assumption is incorrect and causes an overstatement of the variance to the amount of ${}_nm \cdot {}_0m^2 \cdot {}_n\sigma^2 \cdot {}_0\sigma^2$.

7. Application. In Table I is shown the distribution of claims under a certain plan of group sickness and accident insurance. The parameters, (4), for this distribution are

(6) $\quad {}_0m = 3.62, \quad {}_0\sigma^2 = 8.1, \quad {}_0\beta_1 = 14, \quad {}_0(\beta_2 - 3) = 15.$

This distribution is in terms of weeks per claim. The insurance company is interested in the financial cost per claim. A study shows that the distribution of the rate of weekly indemnity to which different classes of employees are entitled has the average parameters,

(7) $\quad {}_1m = 15.25, \quad {}_1\sigma^2 = 16.5, \quad {}_1\beta_1 = 20, \quad {}_1(\beta_2 - 3) = 25.$

Since the moment about zero of the product of independent statistics is equal to the product of the moments, it is permissible to multiply together the corre-

TABLE I

Nearest Duration of Claim in Weeks	Number of Claims per Year per 10,000 Employees
0	197
1	418
2	173
3	109
4	84
5	58
6	45
7	35
8	27
9	24
10	20
11	17
12	14
13	128

sponding parameters of (6) and (7) to obtain the average parameters for the distribution of the financial cost per claim. These are

$${}_2m = 55.2, \quad {}_2\sigma^2 = 134, \quad {}_2\beta_1 = 280, \quad {}_2(\beta_2 - 3) = 375.$$

In order to study the distribution of cost under a group of policies for each of which \$180 in claims is expected, we apply equations (5) to obtain the parameters,

(8) $\quad \sigma^2/m^2 = .74, \quad \beta_1 = 1.6, \quad \beta_2 - 3 = 2.1.$

Since the expected number of claims is

$$P = 180/55.2 = 3.3$$

the probability that there will not be any claims under a policy is

$$h(0) = \frac{1}{0!}(3.3)^0 e^{-3.3} = .037.$$

Adjusting the parameters, (8), to remove the zero claims and choosing the scale so as to express the results as loss ratios gives the parameters,

$$m = 61.6\%, \qquad \sigma = 52.8\%, \qquad \beta_1 = 1.57, \qquad \beta_2 = 4.90.$$

A Pearson Type I curve fitted to these parameters intersects the axis well below the zero point. Therefore β_2 was reduced to 4 59 which gives the expected distribution shown in Table II.

Table II also shows the actual distribution of loss ratios experienced by one of the larger group insurance carriers under policies in this class. The Chi-

TABLE II

Experience under Group Sickness and Accident Insurance Policies

Ratio of Losses to Premiums	Number of Policies	
	Expected	Actual
0	18	11
.01– .09	47	37
.10– .19	53	45
.20– .29	50	56
.30– .39	45	38
.40– .49	41	47
.50– .59	36	39
.60– .69	32	41
.70– .79	28	37
.80– .89	24	20
.90– .99	21	29
1.00–1.19	32	30
1.20–1.39	23	22
1.40–1.59	17	22
1.60–1.99	19	14
2.00 and over	11	9

square test for goodness of fit gives,

$$\chi^2 = 23, \qquad 14 \text{ degrees of freedom,}$$

which corresponds to a probability of 5 per cent. Thus it is apparent that theory and experience are in fair agreement considering that no allowance was made for the lack of homogeneity "between policies." (This should not be confused with the homogeneity "within policies" covered in the theory.)

If the expected number of events is small, especially if the weight distribution is irregular or discrete, it is sometimes advisable to use the following method:

1. Use summation or approximate integration to obtain the distribution, $f(y, n)$, of the sum of n independent weights for $n = 1, 2, 3$, and 4. The formula is

$$f(y, n+1) = \int_0^y f(x) f(y-x, n)\, dx.$$

2. Determine the generalized Poisson distribution for P, the expected number of events, equal to some small number, say $\frac{1}{4}$. The formula is

$$F(y, P) = \Sigma \frac{1}{n!} P^n e^{-P} f(y, n).$$

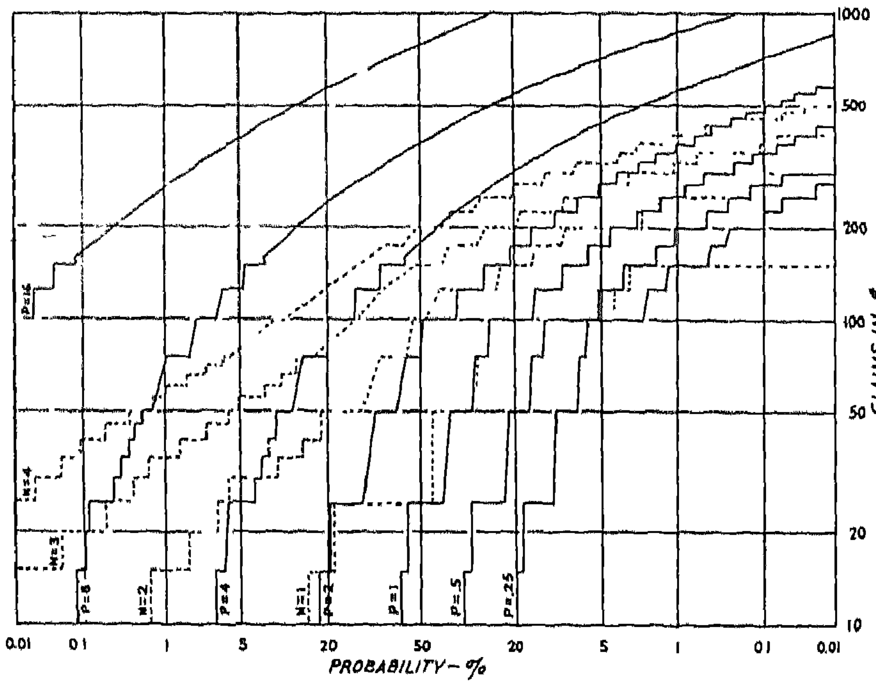


FIG. 1 Surgical Fee Insurance. ----, Distribution, $f(y, n)$, of the sum of n independent claims. —— Distribution, $F(y, P)$, of the sum of the claims when P claims are expected. The average claim is \$50.

Example: If the expected claims under a policy are \$100 ($P = 2$) and if the actual claims are \$490, the probability of an experience as bad as this occurring because of chance factors is 0.1%.

3. Use summation or approximate integration to obtain $F(y, P)$ for $P = \frac{1}{2}$, 1, 2, 4, $\cdots$ by the formula

$$F(y, 2P) = \int_0^y F(x, P) F(y-x, P)\, dx.$$

4. If the calculations are carried on from both tails and if the results are plotted on probability graph paper, it is often possible to fill in the central sec-

tions by interpolation. Such interpolations should be adjusted to reproduce the correct mean This method is illustrated in fig. 1 in the case of surgical fee insurance.

8. Summary. In this paper the Poisson distribution is generalized to allow for the assignment of varying weights to events when the number of events follows the Poisson law. The ability of the Poisson distribution to handle heterogeneous data is preserved in the generalization. An example is given showing that the distribution of certain insurance statistics agrees with that predicted by the theory.

THE CONSTRUCTION OF ORTHOGONAL LATIN SQUARES[1]

By Henry B. Mann[2]

Columbia University

A Latin square is an arrangement of m variables $x_1, x_2, \cdots, x_m$ into m rows and m columns such that no row and no column contains any of the variables twice. Two Latin squares are called orthogonal if when one is superimposed upon the other every ordered pair of variables occurs once in the resulting square.

The rows of a Latin square are permutations of the row $x_1, x_2, \cdots, x_m$. Let P_i be the permutation which transforms $x_1, x_2, \cdots, x_m$ into the ith row of the Latin square. Then $P_i P_j^{-1}$ leaves no variable unchanged for $i \neq j$. For otherwise one column would contain a variable twice. On the other hand each set of m permutations $P_1, P_2, \cdots, P_m$ such that $P_i P_j^{-1}$ leaves no variable unchanged generates a Latin square. We may therefore identify every Latin square with a set of m permutations $(P_1, P_2, \cdots, P_m)$ such that $P_i P_j^{-1}$ leaves no variable unchanged.

Now let $(P_1, P_2, \cdots, P_m)$, $(Q_1, Q_2, \cdots, Q_m)$ be a pair of orthogonal Latin squares. We shall show that $(P_1^{-1}Q_1, P_2^{-1}Q_2, \cdots, P_m^{-1}Q_m)$ is a Latin square. $P_i^{-1}Q_i$ is the transformation which transforms the ith row of $(P_1, P_2, \cdots, P_m)$ into the ith row of $(Q_1, Q_2, \cdots, Q_m)$. Since every pair of variables occurs exactly once if the second square is imposed upon the first, the square $(P_1^{-1}Q_1, P_2^{-1}Q_2, \cdots, P_m^{-1}Q_m)$ contains for every i and k a permutation which transforms x_i into x_k. But then it can not contain two permutations which transform x_i into x_k. This argument can be reversed and it follows that $(P_1, P_2, \cdots, P_m)$ and $(Q_1, Q_2, \cdots, Q_m)$ are orthogonal if and only if $(P_1^{-1}Q_1, P_2^{-1}Q_2, \cdots, P_m^{-1}Q_m)$ is a Latin square.

Denote now by an m sided square S any set of m permutations $(S_1, S_2, \cdots, S_m)$ and by the product SS' of two squares S and S' the square $(S_1S_1', S_2S_1', \cdots, S_mS_m')$. Then we can state: Two Latin squares L_1 and L_2 are orthogonal if and only if there exists a Latin square L_{12} such that

$$L_1 L_{12} = L_2 . \tag{1}$$

Now let $L_1, L_2, \cdots, L_r$ be a set of r mutually orthogonal Latin squares. Then we must have $L_i L_{ik} = L_k$ where L_{ik} is a Latin square if $i \neq k$. Hence we have the theorem

Theorem 1: *The Latin squares $L_1, L_2, \cdots, L_r$ are orthogonal if and only if there exist $r(r-1)$ Latin squares $L_{ik} (i \neq k)$ such that $L_i L_{ik} = L_k$.*

Corollary: *If L^i, L^k and L^{i-k} are Latin squares then L^i is orthogonal to L^k.* For instance if L and L^2 are Latin squares then L is orthogonal to L^2.

---

[1] Presented to the Mathematical Society October 31st, 1942. After I submitted this paper for publication Dr. Edward Fleisher sent me his thesis on Eulerian squares which he submitted in 1934 and in which he proved Theorem 3 in a different manner.

[2] Research under a grant in aid of the Carnegie Corporation of New York.

If $A = (A_1, A_2, \cdots, A_m)$ and P is any permutation then we put $PA = (PA_1, PA_2, \cdots, PA_m)$ and $AP = (A_1P, A_2P, \cdots, A_mP)$. If A is a Latin square then also AP and PA are Latin squares. If A is orthogonal to B then AP is orthogonal to BQ for any permutations P and Q For if $AC = B$ then $AP(P^{-1}CQ) = BQ$, since the associative law holds for the operations indicated. This means that A and B remain orthogonal if we permute the variables in both squares in any arbitrary way.

Hence if A is orthogonal to B also AA_1^{-1} is orthogonal to BB_1^{-1}. We can therefore, while preserving orthogonality, always transform the pair A and B so that $A_1 = B_1 = 1$ where 1 denotes the identity. We shall then say that the pair A, B is written in the reduced form.

DEFINITION 1. *If A is orthogonal to B, and if in the reduced form the permutations of A are the same as those of B in a different order, and if these permutations form a group G, then the pair A and B is said to be based on the group G.*

A pair of orthogonal Latin squares is called a Graeco-Latin square. The Graeco-Latin squares constructed by Bose [1] Stevens [2] and Fisher and Yates [3] are all based on groups. There exist Graeco-Latin squares, however, which are not based on a group.

If the orthogonal pair A, B is based on a group G and if $AC = B$ then also C contains only permutations of G, and since C is a Latin square it must contain all the permutations of G. Calling C_i the image of A_i we obtain a biunique mapping S of G into itself. Let $A_i^S = C_i$ then $B_i = A_iA_i^S$ and S has therefore the property that every element of G is of the form XX^S where X is in G.

DEFINITION 2: *A biunique mapping S of a group G into itself will be called a complete mapping if every element of G can be represented in the form XX^S where X is an element of G and X^S the image of X under the mapping S.*

If an abstract group G of order m admits a complete mapping S then we can immediately construct an m sided Graeco-Latin square based on G. To do this we represent G as a regular permutation group. Let $P_1, P_2, \cdots, P_m$ be the permutations of this representation. Then $A = (P_1, P_2, \cdots, P_m)$, $C = (P_1^S, P_2^S, \cdots, P_m^S)$ and $B = (P_1P_1^S, P_2P_2^S, \cdots, P_mP_m^S)$ are Latin squares and hence A is orthogonal to B and AP_1^{-1} and $B(P_1P_1^S)^{-1}$ form a reduced pair.

If $L_1, L_2, \cdots, L_r$ are orthogonal Latin squares and $L_iL_{ik} = L_k$ then we form the product

$$L_1L_{12}L_{23} \cdots L_{r-1r}. \tag{2}$$

From $L_iL_{ik} = L_k$, $L_kL_{kj} = L_j$ we find $L_iL_{ik}L_{kj} = L_j$ and hence $L_{ik}L_{kj} = L_{ij}$. L_{ik} is therefore orthogonal to L_{ij}. The product (2) has the property that for any $s \leqslant r$ the product of s successive factors is a Latin square On the other hand if a product of r Latin squares $L_1, L_{12}, \cdots, L_{r-1r}$ has this property then the Latin squares $L_1, L_2, \cdots, L_r$ where $L_i = L_1L_{12}L_{23} \cdots L_{i-1i}$ are orthogonal.

DEFINITION 3: *A set of r orthogonal Latin squares will be called based on a group G if every pair in the set is based on G.*

If $L_1, L_2, \cdots, L_r$ are based on a group G then G must admit r mappings $S_1 = 1, S_2, \cdots, S_r$ into itself such that every element of G can be written in

the form $X^{S_i+S_{i+1}+\cdots+S_{i+h}}$ for every i and h with $1 \leqslant i \leqslant r$ and $0 \leqslant h \leqslant r - i$, where $A^{S+S'} = A^S A^{S'}$, and A^S is the image of A under the mapping S.

DEFINITION 4: *The mappings $S_1 = 1, S_2, \cdots, S_r$ of a group G into itself will be called r-fold complete if every element of G is of the form $X^{S_i+S_{i+1}+\cdots+S_{i+h}}$ for every i and h with $1 \leqslant i \leqslant r$ and $0 \leqslant h \leqslant r - i$.*

Now let G be an abstract group of order m admitting an r-fold complete set of mappings $S_1 = 1, S_2, \cdots, S_r$. Put

$$L_i = (1^{S_1+S_2+\cdots+S_i}, P_2^{S_1+S_2+\cdots+S_i}, \cdots, P_m^{S_1+S_2+\cdots+S_i})$$

where $1, P_2, \cdots, P_m$ is a regular representation of G. Then $L_1, L_2, \cdots, L_r$ is a set of r orthogonal Latin squares based on G. Put $A_i = 1^{S_1+S_2+\cdots+S_i}$ then $L_1A_1^{-1}, \cdots, L_rA_r^{-1}$ are written in the reduced form. Hence we have

THEOREM 2: *A set of r orthogonal Latin squares based on a group G exists if and only if G admits an r-fold complete set of mappings.*

If G is of order $m = 4n + 2 = 2m'$ then G has a self-conjugate subgroup H of order m'. Suppose G admits a complete mapping S. We have

$$G = H + HA.$$

$XX^S \subset H$ if either X and X^S or neither of them are in H. Further $XX^S \subset HA$ if either X or X^S but not both of them are in H.

Let a be the number of elements $X \subset H$ such that $X^S \subset H$,
b the number of elements $X \subset H$ such that $X^S \subset HA$,
c the number of elements $X \subset HA$ such that $X^S \subset H$,
then $a + b = m'$, $a + c = m'$. Of the products XX^S exactly $b + c$ are in HA. Hence $b + c = m'$, $a = b$ and therefore $m' = 2a$, which is impossible since m' is odd. We have therefore:

THEOREM 3: *No $4n + 2$ — sided Graeco-Latin square based on a group can exist.*

If a group G admits r automorphisms $T_1 = 1, T_2, \cdots, T_r$ such that $X^{T_i} \neq X^{T_j}$ for $i \neq j$ and $X \neq 1$ then the mappings $S_1 = 1$, $S_i = X^{-T_{i-1}}X^{T_i}$ for $i = 2, 3, \cdots, r$ are r-fold complete; for if

$$X^{S_i+S_{i+1}+\cdots+S_{i+h}} = Y^{S_i+S_{i+1}+\cdots+S_{i+h}}$$

we have for $i = 1$

$$X^{T_{1+h}} = Y^{T_{1+h}}$$

and for $i > 1$

$$X^{-T_{i-1}}X^{T_{i+h}} = Y^{-T_{i-1}}Y^{T_{i+h}}$$

and therefore

$$(YX^{-1})^{T_{i-1}} = (YX^{-1})^{T_{i+h}}$$

and hence $Y = X$ in both cases since by hypothesis $X^{T_i} \neq X^{T_j}$ for $i \neq j$ and $X \neq 1$, $X^{S_i+\cdots+S_{i+h}}$ therefore takes m different values and reproduces every element of G.

If we represent G as a regular permutation group then the squares $L_1 = (1, P_2, \cdots, P_m)$, $L_2 = (1, P_2^{T_2}, \cdots, P_m^{T_2}), \cdots, L_r = (1, P_2^{T_r}, \cdots, P_m^{T_r})$ are orthogonal Latin squares by Theorems 1 and 2. There exist however complete

mappings which are not derivable from automorphisms. For instance every group of odd order admits the complete mapping $A^S = A$ but $A^T = A^2$ is not an automorphism if the group is not abelian

Most of the sets of orthogonal Latin squares that have been constructed so far are based on abelian groups of type $(p, p, \cdots, p)$ and the mappings of the squares of the sets into each other are automorphisms of this group R. C. Bose [1] and W. L. Stevens [2] for instance use the cyclic group of automorphisms of the additive group of a G. F. (p^n) induced through multiplication by the elements of the Galois field that are different from 0 In this way they assure that different automorphisms will map the same element into different elements. They give a convenient method for finding a base element of the group of automorphisms. In this way they reduce considerably the labor involved in the construction of $p^n - 1$ orthogonal Latin squares of side p^n. The 9 x 9 squares in the statistical tables by Fisher and Yates [3] are also based on the abelian group of type (3,3) but another set of automorphisms is used.

If $m = p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n}$ (p_i prime $p_i \neq p_k$ for $i \neq k$) and if $r = \min p_i^{e_i} - 1$ then a set of r orthogonal Latin squares can always be constructed from the abelian group of type $(p_1 \cdots p_1, p_2 \cdots p_2, \cdots, p_n, \cdots, p_n)$ and its automorphisms. This can be done by finding r automorphisms $T_1^{(i)}, T_2^{(i)}, \cdots, T_r^{(i)}$ for each of the subgroups of order $p_i^{e_i}$ such that $T_k^{(i)} T_j^{(i)-1}$ leaves no element unchanged except 1. If we apply the automorphisms $T_j^{(1)}, T_j^{(2)}, \cdots, T_j^{(n)}$ simultaneously, for $j = 1, 2, \cdots, r$, we obtain r automorphisms of the desired type.

Once the automorphisms are known the construction of the set of orthogonal Latin squares can easily be carried out. To do this we have to write down the multiplication table of the group and obtain the orthogonal squares by interchanging the rows in accord with the automorphisms

DEFINITION 5: *A set of orthogonal Latin squares derived from a group and its automorphisms will be called constructed by the automorphism method.*

We now prove:

THEOREM 4. *Let c_q be the number of classes of elements of order q of a group G. Let $s = \min c_q$; then not more than s orthogonal Latin squares can be constructed from G by the automorphism method.*

PROOF: Let T be an automorphism which leaves no element unchanged except 1. If A is of order q then A^T is also of order q If $A^T = P^{-1}AP$ then there exists an element Q such that $P = Q^{-1}Q^T$ because, as we have shown, every element can be represented in the form $X^{-1}X^T$. But then

$$(QAQ^{-1})^T = QPP^{-1}APP^{-1}Q = QAQ^{-1}.$$

Hence $A = 1$. T can therefore not transform any element except 1 into an element of the same class. Hence not more than $s = \min c_q$ automorphisms, $T_1, \cdots, T_s$ can exist such that $T_i^{-1}T_j$ leave no element except 1 fixed and this proves our theorem

COROLLARY. *If $m = p_1^{e_1}p_2^{e_2} \cdots p_n^{e_n}$ (p_i prime $p_j \neq p_k$ for $j \neq k$) then not more than $r = \min p_i^{e_i} - 1$ orthogonal m-sided Latin squares can be constructed from any group with the automorphism method.*

PROOF: The Sylow group of order $p_i^{e_i}$ contains a representative of every class of elements of order p_i hence $\min c_q \leqslant \min p_i^{e_i} - 1$.

Below are given two examples of Graeco-Latin squares obtained from complete mappings which are not obtained from automorphisms. Neither could have been obtained by combining Graeco-Latin squares constructed by the method of Bose [1] and Stevens [2].

The first example is based on the abelian group of type (2,2,3). If the basis elements are defined by $P^2 = R^2 = Q^3 = 1$ the complete mapping used is given by

$$L_1 = (1, P, R, PR, Q, PQ, RQ, PRQ, Q^2, PQ^2, RQ^2, PRQ^2)$$

$$L_{12} = (1, RQ, PRQ^2, PQ^2, Q, RQ^2, PR, P, Q^2, R, PRQ, PQ)$$

$$L_2 = (1, PRQ, PQ^2, RQ^2, Q^2, PR, PQ, RQ, Q, PRQ^2, P, R).$$

The second square is based on the regular representation of the A_4 the alternating group in 4 variables. The generating relations are $P^2 = R^2 = Q^3 = 1$, $QP = RQ$, $QR = PRQ$. The complete mapping is given by

$$L_1 = (1, P, R, PR, Q, PQ, RQ, PRQ, Q^2, PQ^2, RQ^2, PRQ^2)$$

$$L_{12} = (1, R, PR, P, Q, PQ, RQ, PRQ, Q^2, PQ^2, RQ^2, PRQ^2)$$

$$L_2 = (1, PR, P, R, Q^2, PRQ^2, PQ^2, RQ^2, Q, RQ, PRQ, PQ).$$

EXAMPLE 1

1,1	2,2	3,3	4,4	5,5	6,6	7,7	8,8	9,9	10,10	11,11	12,12
2,8	1,7	4,6	3,5	6,12	5,11	8,10	7,9	10,4	9,3	12,2	11,1
3,10	4,9	1,12	2,11	7,2	8,1	5,4	6,3	11,6	12,5	9,8	10,7
4,11	3,12	2,9	1,10	8,3	7,4	6,1	5,2	12,7	11,8	10,5	9,6
5,9	6,10	7,11	8,12	9,1	10,2	11,3	12,4	1,5	2,6	3,7	4,8
6,4	5,3	8,2	7,1	10,8	9,7	12,6	11,5	2,12	1,11	4,10	3,9
7,6	8,5	5,8	6,7	11,10	12,9	9,12	10,11	3,2	4,1	1,4	2,3
8,7	7,8	6,5	5,6	12,11	11,12	10,9	9,10	4,3	3,4	2,1	1,2
9,5	10,6	11,7	12,8	1,9	2,10	3,11	4,12	5,1	6,2	7,3	8,4
10,12	9,11	12,10	11,9	2,4	1,3	4,2	3,1	6,8	5,7	8,6	7,5
11,2	12,1	9,4	10,3	3,6	4,5	1,8	2,7	7,10	8,9	5,12	6,11
12,3	11,4	10,1	9,2	4,7	3,8	2,5	1,6	8,11	7,12	6,9	5,10

EXAMPLE 2

1,1	2,2	3,3	4,4	5,5	6,6	7,7	8,8	9,9	10,10	11,11	12,12
2,4	1,3	4,2	3,1	6,8	5,7	8,6	7,5	10,12	9,11	12,10	11,9
3,2	4,1	1,4	2,3	7,6	8,5	5,8	6,7	11,10	12,9	9,12	10,11
4,3	3,4	2,1	1,2	8,7	7,8	6,5	5,6	12,11	11,12	10,9	9,10
5,9	7,12	8,10	6,11	9,1	11,4	12,2	10,3	1,5	3,8	4,6	2,7
6,12	8,9	7,11	5,10	10,4	12,1	11,3	9,2	2,8	4,5	3,7	1,6
7,10	5,11	6,9	8,12	11,2	9,3	10,1	12,4	3,6	1,7	2,5	4,8
8,11	6,10	5,12	7,9	12,3	10,2	9,4	11,1	4,7	2,6	1,8	3,5
9,5	12,7	10,8	11,6	1,9	4,11	2,12	3,10	5,1	8,3	6,4	7,2
10,7	11,5	9,6	12,8	2,11	3,9	1,10	4,12	6,3	7,1	5,2	8,4
11,8	10,6	12,5	9,7	3,12	2,10	4,9	1,11	7,4	6,2	8,1	5,3
12,6	9,8	11,7	10,5	4,10	1,12	3,11	2,9	8,2	5,4	7,3	6,1

REFERENCES

[1] R. C. Bose, "On the application of the properties of Galois fields to the problem of construction of Hyper-Graeco-Latin-squares," *Sankhya*, 1938

[2] W L. Stevens, "The completely orthogonalized Latin-square," *Annals of Eugenics*, 1939.

[3] R. A. Fisher and F Yates, *Statistical Tables for Agricultural, Biological, and Medical Research*, Edinburgh. Oliver and Boyd

A METHOD OF DETERMINING EXPLICITLY THE COEFFICIENTS OF THE CHARACTERISTIC EQUATION

By P. A. Samuelson

Massachusetts Institute of Technology

1. Introduction. When an investigator is interested in all of the latent roots of the characteristic equation of a matrix and not in its latent vectors, it is sometimes desirable to expand out the determinental equation in order to determine explicitly the polynomial coefficients ($p_1, p_2, \cdots, p_n$) in the expression

(1) $$D(\lambda) = | \lambda I - a | = \lambda^n + p_1\lambda^{n-1} + \cdots + p_{n-1}\lambda + p_n .$$

This can be done in a variety of ways, all of which are necessarily somewhat tedious for high order matrices. Except for sign the coefficients are respectively the sum of a's principal minors of a given order. These can be computed efficiently by "pivotal" methods [1]. Alternatively through the utilization of the Cayley-Hamilton theorem, whereby a matrix satisfies its own characteristic equation, the p's appear as the solution of n linear equations [2, 3]. In a third method Horst has employed Newton's formula concerning the powers of roots to derive the p's as the solution of a triangular set of equations, the coefficients of the latter only being attained after considerable matrix multiplication [4]. A fourth method suggested to me by Professor E. Bright Wilson, Jr. of Harvard University, consists of evaluating $D(\lambda)$ for n values of λ, presumably by efficient "Doolittle" methods; to these n points, Lagrange's interpolation formula is applied to determine the n coefficients explicitly.

2. The New Method. The present paper describes a new computational method based upon well-known dynamical considerations. A single nth order differential equation can be converted into "normal" form, involving n first order differential equations. This is easily done by defining appropriate new variables. If the original nth order differential equation is written as

(2) $$X^{(n)}(t) + p_1X^{(n-1)}(t) + \cdots + p_{n-1}X'(t) + p_n = 0,$$

then the new normal system can be written as

(3) $$X_i'(t) = \sum_1^n b_{ij} X_j(t), \qquad (i = 1, \cdots n)$$

where

(4) $$[b_{ij}] = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -p_n & -p_{n-1} & -p_{n-2} & \cdots & -p_1 \end{bmatrix}$$

is the so-called companion matrix to the polynomial in question.

The reverse process of going from a normal system in many variables to a single high order equation is not so simple. Yet it can be done, and in so doing we attain the required polynomial coefficients [5]. If

$$x'(t) = ax(t) \tag{5}$$

represents the normal system in matrix form, then symbolically

$$D\left(\frac{d}{dt}\right) X_1(t) = X_1^{(n)}(t) + p_1 X_1^{(n-1)}(t) + \cdots + p_{n-1} X_1'(t) + p_n . \tag{6}$$

Because we wish to find out the expanded form of $D(\lambda)$, this relationship is of no use to us. Since similar matrices have the same characteristic equation, ours is the problem of finding a non-singular matrix C, such that

$$C^{-1}aC = b, \tag{7}$$

where b is of the form given in equation (4).

This problem can be approached from an elementary algebraic viewpoint. The relationships in (5) represent n linear equations between $2n$ variables, $[X_1(t), X_2(t), \cdots, X_n(t), X_1'(t), X_2'(t), \cdots, X_n'(t)]$. These are not sufficient to eliminate the $2(n-1)$ variables not involving the subscript 1. However, inasmuch as (5) holds for all values of t we may differentiate it repeatedly until we finally have the system of equations

$$\begin{array}{lr}
-X_1^{(n)} + \cdots + a_{11}X_1^{(n-1)} + \cdots + a_{1n}X_n^{(n-1)} & = 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\quad -X_n^{(n)} + a_{n1}X_1^{(n-1)} + \cdots + a_{nn}X_n^{(n-1)} & = 0 \\
\qquad -X_1^{(n-1)} + \quad + a_{11}X_1^{(n-2)} + \cdots + a_{1n}X_n^{(n-2)} & = 0 \\
\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\qquad\quad -X_n^{(n-1)} + a_{n1}X_1^{(n-2)} + \cdots + a_{nn}X_n^{(n-2)} & = 0 \\
\qquad\qquad -X_1' + \cdots + a_{11}X_1 + \cdots + a_{1n}X_n & = 0 \\
\qquad\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\qquad\qquad -X_n' + a_{n1}X_1 + \cdots + a_{nn}X_n & = 0
\end{array} \tag{8}$$

These are n^2 linear equations in $n^2 + n$ variables. We wish to eliminate all variables which have a subscript other than one; namely, $(X_2, \cdots, X_n, X_2', \cdots, X_n', \cdots, X_2^{(n)}, \cdots X_n^{(n)})$. These are $(n+1)(n-1) = n^2 - 1$ in number. We may utilize all but one of the n^2 equations to perform this elimination. The remaining equation after substitution will be the desired high order equation, and its coefficients are the polynomial coefficients.

Ordinarily one would solve all but one of the equations for the values of the variables to be eliminated. These would then be substituted into the remaining equation. Actually from the computational standpoint it is unnecessary to solve completely for any unknowns. The so-called "forward" solution of the usual Gauss-Doolittle technique automatically performs the elimination or

substitution, without necessary recourse to a "back" solution for the values of the eliminated variables. These values are in any case of no interest.

There is no unique order in which the equations must be reduced. Indeed, when one order fails because a leading principal minor vanishes, we may switch to another. A suggested convenient order is given below. Let

$$\left[\begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array}\right] = \left[\begin{array}{c|c} a_{11} & R \\ \hline S & M \end{array}\right]; \qquad I = (\delta_{ij}), \qquad (i, j = 1, \cdots, n-1)$$

Then, consider the partitioned matrix

$$(9) \qquad W = \left[\begin{array}{cccccc|ccccc} -I & M & 0 & \cdots & 0 & 0 & 0 & -S & 0 & \cdots & 0 \\ 0 & -I & M & \cdots & 0 & 0 & 0 & 0 & -S & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -I & M & 0 & 0 & 0 & \cdots & -S \\ 0 & 0 & 0 & \cdots & 0 & R & 0 & 0 & 0 & \cdots & -a_{11} \\ 0 & 0 & 0 & \cdots & R & 0 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & R & \cdots & 0 & 0 & 0 & 1 & -a_{11} & \cdots & 0 \\ \hline 0 & R & 0 & \cdots & 0 & 0 & 1 & -a_{11} & 0 & \cdots & 0 \end{array}\right]$$

It is simply the matrix of the equations in (8) with the variables $(X_1, X_1', \cdots, X_1^{(n)})$ shifted over to the right-hand side, and with the equations in which the variable one leads off being placed at the bottom.

If the usual "forward" Doolittle technique is followed, then the final elements computed, corresponding to the elements in the lower right-hand box, are the coefficients $(1, p_1, p_2, \cdots, p_n)$. It is the present writer's experience that the Crout form [6], like Dwyer's [7] the last word in Doolittle abbreviation, is to be recommended, particularly since we are dealing with an asymmetrical matrix. A clerk masters its ritual in a few minutes, and the speeds achieved once the operations become mechanical are impressive.

For the trivial case of determining the coefficients corresponding to a two by two matrix the W matrix is of the form

$$(10) \qquad \left[\begin{array}{ccc|ccc} -1 & a_{22} & 0 & 0 & -a_{21} & 0 \\ 0 & -1 & a_{22} & 0 & 0 & -a_{21} \\ 0 & 0 & a_{12} & 0 & 1 & -a_{11} \\ \hline 0 & a_{12} & 0 & 1 & -a_{11} & 0 \end{array}\right]$$

The Auxiliary Crout matrix becomes

$$(11) \qquad \left[\begin{array}{ccc|ccc} -1 & a_{22} & 0 & 0 & -a_{21} & 0 \\ 0 & -1 & a_{22} & 0 & 0 & -a_{21} \\ 0 & 0 & a_{12} & 0 & 1 & -a_{11} \\ \hline 0 & -a_{12} & a_{22} & 1 & (-a_{11} - a_{22}) & (-a_{12}a_{21} + a_{11}a_{22}) \end{array}\right]$$

The answer in the lower right-hand box will immediately be recognized as the correct one. I have found it convenient to vary the precise Crout routine by dividing vertical columns by the "leading" diagonal element, rather than horizontal columns This is a matter of indifference and saves some computations. As in the higher order cases, the presence of the identity matrix along the diagonal reduces most of the computations to mere copying. Actually the intelligent computer will soon notice that most of the copying may be eliminated since the numbers in question are to be added in later in other sums of products. After eliminating unknowns corresponding to the equations above the line on which (9) is written, there results the system

$$
(12)\quad \left[\begin{array}{l|ccccccccc}
R & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -a_{11} \\
RM & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -a_{11} & -RS \\
RM^2 & 0 & 0 & 0 & 0 & \cdots & 1 & -a_{11} & -RS & -RMS \\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & & \\
RM^{n-1} & 1 & -a_{11} & -RS & -RMS & \cdots & & \cdots & \cdots & -RM^{n-2}S
\end{array}\right]
$$

Thus, it would be simpler to start from this stage, avoiding unnecessary copying.

This remark shows that the present method is related to the Cayley-Hamilton methods described in [2] and [3], since the above set is derivable from the set

$$
(13)\quad \left[\begin{array}{ll|ccccc}
e_1' & A^0 & 1 & 0 & 0 & \cdots & 0 \\
e_1' & A^1 & 0 & 1 & 0 & \cdots & 0 \\
e_1' & A^2 & 0 & 0 & 1 & \cdots & 0 \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\cdot & & & & & & \\
e_1' & A^n & 0 & 0 & 0 & \cdots & 1
\end{array}\right]
$$

The last named set appears in the Cayley-Hamilton method when the first row of the powers of the original matrix are used in setting up n equations to determine our n unknowns. Although related, the two methods are distinct since in the Cayley-Hamilton method one would arrive at a different set of equations after straightforward elimination of one variable, and since it would be shorter to dispense with the identity matrix used in the Aitken method in favor of the solution of a single set of equations by the usual Doolittle "back-solution."

The reader will easily see how the method may be modified to handle the more general case of determining the coefficients of

$$
(14)\qquad D(\lambda) = |c\lambda + a| = 0,
$$

where c and a are any matrices. The method also can be used to reduce a polynomial equation involving a determinant of the nth order, each of whose coefficients are of a given degree in λ, to a lower order determinant whose coefficients are of higher degree in λ.

The present method derives the p's as the algebraic solution of high order linear equations. It would therefore seem inferior to those methods which need only solve a system of n equations. However, two remarks are in order. The matrix of the high order system can be written down immediately without computation. Furthermore, most of the elements in the matrix are zeros, so that a mere counting of the equations is not a true indication of the labor involved.

3. Some comparsions between present method and other methods. Within the brief compass of the present work it is not possible to give an exhaustive appraisal of the comparative computational efficiencies of the methods mentioned. In general, a computing method is to be judged in terms of the number of multiplications that it involves, although other considerations such as the number of additions, the magnitude and sign of the numbers handled, the repetitiveness of the operations involved, the adaptability to punch card machinery, etc. are modifying factors. In this discussion the *power* of a method will be taken to be an inverse function of the number of multiplications that it involves.

It may be said first of all that inasmuch as the minimum number of multiplications involved in computing an nth order determinant is of the order of n^3, even with the most efficient "pivotal" methods, direct computation of the coefficients by principal minors involves, for sufficiently large n, computation of the order of n^4. The same is true of the Wilson method described above. The Horst method, and any other that requires the explicit n powers of an nth order matrix, also asymptotically requires multiplications of the order of n^4. This does not mean that the above three methods are equally powerful for small n, nor even asymptotically, since the coefficients of the n^4 term in the formula for the requisite number of multiplications may not be equal. In fact, Riersol [1] has shown that his method is better than Horst's for small n, but asymptotically less powerful.

It can also be shown that the Cayley-Hamilton methods which simply involve products of the powers of a matrix with row or column vectors are asymptotically more powerful than any of the above methods, the work only increasing as the cube of n. This is true whether the longer Aitken form of reduction is employed or whether the usual Doolittle back-solution is followed. The present method is also an efficient one in the sense that its requisite number of multiplications increases with the cube of n. For small values of n and asymptotically it can be shown to be more powerful than the Cayley-Hamilton method which uses the Aitken method of reduction, although in the limit as n becomes large the ratio of the powers of the two methods approaches unity.

It is of the greatest interest to compare the power of the new method with the shorter Doolittle *C-H* method. It can easily be shown that the coefficients of n^3 in the expressions giving the respective requisite number of multiplications differ in such a way as to make the *C-H* method more powerful after some value of n,

the ratio of the respective powers approaching the limit 8/9 However, for low order matrices the new method is the more powerful The reader may easily verify this for the case of a second order matrix. Below a sixth order matrix the present method seems to involve the smaller number of multiplications. For a sixth order matrix the two methods seem to involve the same number of multiplications (multiplications by unity not being counted). For matrices of the seventh order or higher the *C-H* method seems to be optimal.

As compared to an explicit evaluation of the coefficients by a straightforward computation of principal minors according to the fundamental definition of a determinant as the sum of signed products of elements, all of the methods discussed are efficient, since the work in the former increases faster than any power of n. However, for each of the methods discussed, in singular cases the method of reduction may fail so that modified procedures will be necessary. In actual practice such singularities will "almost never" be encountered But in the neighborhood of such singular points the computations become extremely sensitive to any rounding off of digits. Consequently, it is from the nature of the case impossible ever to develop exact rules for the maximum error involved in any given calculation.

REFERENCES

[1] O. Riersol, "Recurrent Computations of all Principal Minors of a Determinant," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 193–198.

[2] R. A. Fraser, W. J. Duncan, and A. R. Collar, *Elementary Matrices*, pp. 141–142.

[3] M. M. Flood, "A Computational Procedure for the Method of Principal Components," *Psychometrika*, Vol. 5 (1940), pp. 169–172.

[4] P. Horst, "A Method for Determining the Coefficients of a Characteristic Equation," *Annals of Math. Stat.*, Vol. 5 (1935), pp. 83–84.

[5] F. R. Moulton, *Differential Equations*, pp. 6–9

[6] P. D. Crout, "A Short Method for Evaluating Determinants and Solving Systems of Linear Equations with Real or Complex Coefficients," *American Institute of Electrical Engineers*, Vol. 60 (1941).

[7] P. S. Dwyer, "The Solution of Simultaneous Equations," *Psychometrika*, Vol. 6 (1941), pp. 101–129

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A NOTE ON THE THEORY OF MOMENT GENERATING FUNCTIONS

By J. H. Curtiss

Cornell University

Let X be a one-dimensional variate and let $F(x)$ be its distribution function.[1] The function

$$G(\alpha) = E(e^{\alpha X}) = \int_{-\infty}^{+\infty} e^{\alpha x}\, dF(x), \qquad \alpha \text{ real},$$

in which the integral is assumed to converge for α in some neighborhood of the origin, is called the moment generating function of X. In dealing with certain distribution problems, this function has been widely used by statisticians, and especially by the English writers, in place of the closely-related characteristic function $f(t) = E(e^{itX})$. It is known that a characteristic function uniquely determines the corresponding distribution, and that if a sequence of characteristic functions approaches a limit, the corresponding sequence of distribution functions does likewise. (These results are more accurately stated below.) The appropriate analogues for the moment generating function of these theorems are apparently not too readily accessible in the literature, if they have been treated at all, and it seems worthwhile to record them in this note.

Henceforth we abbreviate distribution function to d.f., moment generating function to m.g.f., and characteristic function to c.f. The variables α and t will always be real, in contradistinction to the complex variable s, to be introduced in the next paragraph.

The uniqueness property of the c f. may be stated as follows: If $F_1(x)$ and $f_1(t)$ are the d.f. and c.f. of one variate, and $F_2(x)$ and $f_2(t)$ are those of another, and if $f_1(t) \equiv f_2(t)$ for all[2] t, then $F_1(x) \equiv F_2(x)$ for all x [1, p 28]. To study the corresponding situation for the m.g.f., we first observe that

$$\varphi(s) = E(e^{sX}) = \int_{-\infty}^{+\infty} e^{sx}\, dF(x), \qquad s \text{ complex},$$

[1] Or cumulative frequency function; our notation and terminology are uniform with that of [1] except for the use of the term "variate" instead of "random variable."

[2] It is possible for two non-identical distributions to have c.f.'s which are identical throughout an interval of values of t containing the origin; an example is given in [4], p 190. The author is obliged to Professor Wintner and Professor Feller for pointing out the existence of this particular example

is a bilateral Laplace-Stieltjes transform. If such a transform exists for real values of s in an interval $-\alpha_1 < s < \alpha_1$, $\alpha_1 > 0$, it must exist for all complex values of s in the strip $-\alpha_1 < \Re s < \alpha_1$, and represent there an analytic function of s [5, p. 238]. Evidently $\varphi(\alpha) = G(\alpha)$, $\varphi(it) = f(t)$. Suppose now that $F_1(x)$, $G_1(\alpha)$, $f_1(t)$, are the d.f., m.g.f., and c.f. of a variate X_1, and $F_2(x)$, $G_2(\alpha)$, $f_2(t)$, are those of X_2. Let $\varphi_1(s) = E(e^{sX_1})$, $\varphi_2(s) = E(e^{sX_2})$, s complex If $G_1(\alpha) \equiv G_2(\alpha)$ for all α in some interval, however small, containing the origin, then by a familiar property of analytic functions [2, p. 116], $\varphi_1(s) \equiv \varphi_2(s)$ throughout the corresponding strip of analyticity, and so on the axis of imaginaries. This means that $f_1(t) \equiv f_2(t)$, all t, and therefore $F_1(x) \equiv F_2(x)$. We have:

THEOREM 1. *A m.g.f. existing in some neighborhood of $\alpha = 0$ uniquely determines the corresponding distribution.*

We turn now to distributions of variable form. Because certain of the versions to be found in the literature are incomplete, it seems worth while to give here a full statement of the basic limit theorem for sequences of c.f.'s, due to P. Lévy and sometimes called Lévy's Continuity Theorem [4, pp. 48–50].

THEOREM 2. *Let the distribution of a variate X_n depend on a parameter n, and let $F_n(x)$ and $f_n(t)$ be the d.f. and c.f. of X_n.*

(a) *If there exists a variate X with d.f. $F(x)$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ at every continuity point of $F(x)$, then $\lim_{n\to\infty} f_n(t) = f(t)$ uniformly in each finite interval on the t-axis, where $f(t)$ is the c.f. of X.*

(b) *If there exists a function $f(t)$ such that $\lim_{n\to\infty} f_n(t) = f(t)$, all t,*[3] *and uniformly*[4] *in some open interval containing the origin, then there exists a variate X with d.f. $F(x)$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ at each continuity point and uniformly in any finite or infinite interval of continuity of $F(x)$. The c.f. of X is $f(t)$, and $\lim_{n\to\infty} f_n(t) = f(t)$ uniformly in each finite interval.*

We now develop the corresponding theorem for the m.g.f In the first place, it is not difficult to see that part (a) will have no direct analogue, even if we add to the hypothesis the conditions that the m.g.f. of X_n exists in some fixed interval for all n and that the m.g.f. of X also exists in some interval. For example, the d.f.

$$F_n(x) = \begin{cases} 0, \; x < -n \\ \frac{1}{2} + k_n \operatorname{arc\,tan} nx, & -n \leqq x < n \\ 1, \; x \geqq n \end{cases}$$

[3] The condition that $\lim_{n\to\infty} f_n(t)$ exist on at least an everywhere dense set of points on the t-axis is essential to the proof as given in Cramer's book [1, pp. 29–30], but is omitted in his statement of the theorem, and is not stated clearly in certain other treatments by other authors.

[4] For a discussion of this uniformity condition, and possible alternatives, see [1, p. 29 (footnote)]. The condition may, for instance, be replaced by the assumption that $f(t)$ is continuous at $t = 0$.

where $k_n = 1/(2 \operatorname{arc} \tan n^2)$, clearly tends as $n \to \infty$ to the d.f.

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqq 0 \end{cases}$$

at all points of continuity of the latter d.f. The m.g.f. corresponding to $F_n(x)$ is

$$G_n(\alpha) = \int_{-n}^{n} k_n e^{\alpha x} \frac{n}{1 + n^2x^2}\, dx,$$

which for each n exists for all α, and the m g.f. corresponding to $F(x)$ is simply the constant 1. Clearly

$$G_n(\alpha) > \int_0^n k_n \frac{|\alpha|^3 x^3}{3!} \cdot \frac{n}{1 + n^2x^2}\, dx,$$

and from this it can easily be verified that $\lim_{n\to\infty} G_n(\alpha) = \infty$ if $\alpha \neq 0$. In short, mere convergence of a sequence of d.f.'s tells little about the behavior of the corresponding sequence of m.g.f.'s.

Part (b) assumes the following form:

THEOREM 3. *Let $F_n(x)$ and $G_n(\alpha)$ be respectively the d.f. and m g.f. of a variate X_n. If $G_n(\alpha)$ exists for $|\alpha| < \alpha_1$ and for all $n \geqq n_0$, and if there exists a finite-valued function $G(\alpha)$ defined for $|\alpha| \leqq \alpha_2 < \alpha_1$, $\alpha_2 > 0$, such that* $\lim_{n\to\infty} G_n(\alpha) = G(\alpha)$, $|\alpha| \leqq \alpha_2$, *then there exists a variate X with d.f. $F(x)$ such that* $\lim_{n\to\infty} F_n(x) = F(x)$ *at each continuity point and uniformly in each finite or infinite interval of continuity of $F(x)$. The m.g.f. of X exists for $|\alpha| \leqq \alpha_2$ and is equal to $G(\alpha)$ in that interval.*

To prove the theorem, we introduce the Laplace transform $\varphi_n(s) = E(e^{sX_n})$ and observe that $|\varphi_n(s)| \leqq \varphi_n(\alpha) = G_n(\alpha)$, $s = \alpha + it$, $n \geqq n_0$, for any s in the strip $-\alpha_1 < \Re s < \alpha_1$. By applying Leibniz's rule for differentiation under an integral sign (extended to Stieltjes integrals), we find [5, p. 240] that

$$G_n''(\alpha) = \int_{-\infty}^{+\infty} x^2 e^{\alpha x}\, dF_n(x), \qquad |\alpha| < \alpha_1,$$

from which it appears that $G_n''(\alpha) > 0$, $|\alpha| < \alpha_1$. This means that the function $G_n(\alpha)$ assumes its maximum value in the interval $|\alpha| \leqq \alpha_2$ at either or both endpoints of the interval. But of course $G_n(\alpha_2)$ and $G_n(-\alpha_2)$ both approach finite limits as n becomes infinite, so it follows that the sequence $\{G_n(\alpha)\}$, $n \geqq n_0$, is uniformly bounded in the interval $|\alpha| \leqq \alpha_2$. Thus the sequence $\{|\varphi_n(s)|\}$, $n \geqq n_0$, is uniformly bounded in the strip $-\alpha_2 \leqq \Re s \leqq \alpha_2$, and moreover has a limit at each point of an infinite set possessing a limit point in the strip (i.e., at each point of the interval $-\alpha_2 \leqq s \leqq \alpha_2$). So by Vitali's Theorem [3, pp. 156–160, 240], there exists an analytic function $\varphi^*(s)$ such that $\lim_{n\to\infty} \varphi_n(s) = \varphi^*(s)$ uniformly in each bounded closed subregion of the strip $-\alpha_2 < \Re s < \alpha_2$. Since $\varphi_n(it)$ is the c.f of X_n, the existence of the limiting distribution follows from Theorem 2(b).

Of course, $\varphi^*(\alpha) = G(\alpha)$, $-\alpha_2 < \alpha < \alpha_2$. It remains to show that $\varphi^*(\alpha)$ is the m.g.f. of X. Theorem 2(b) states that $\varphi^*(it)$ is the c.f. of X. If we can show that the function $\varphi(s) = E(e^{sX})$ exists at least in the strip $-\alpha_2 < \Re s < \alpha_2$, then since $\varphi(s) \equiv \varphi^*(s)$ on the axis of imaginaries, the equality must be valid in the entire strip, and so in particular on the interval of the real axis inside the strip.

It will suffice for this purpose to show that $\varphi(\alpha)$ exists for $-\alpha_2 \leqq \alpha \leqq \alpha_2$. Suppose indeed that $\varphi(\alpha)$ does not exist at some point $\alpha = \alpha_3$ in this interval. That means that if

$$M = [\text{l.u.b. } G_n(\alpha_3), n \geqq n_0],$$

we can find a real number A such that

$$\int_{-A}^{A} e^{\alpha_3 x}\, dF(x) > M. \tag{1}$$

But

$$\int_{-A}^{A} e^{\alpha_3 x}\, dF(x) = \int_{-A}^{A} e^{\alpha_3 x}\, dF_n(x) + \left[\int_{-A}^{A} e^{\alpha_3 x}\, dF(x) - \int_{-A}^{A} e^{\alpha_3 x}\, dF_n(x)\right].$$

Since $\lim_{n\to\infty} F_n(x) = F(x)$ at all continuity points of $F(x)$, and so on an everywhere dense set of points, the Helly-Bray Theorem [5, p. 31] states that the expression in brackets in (2) approaches zero as n becomes infinite. Meanwhile

$$\int_{-A}^{A} e^{\alpha_3 x}\, dF_n(x) \leqq \int_{-\infty}^{+\infty} e^{\alpha_3 x}\, dF_n(x) \leqq M, \qquad n \geqq n_0.$$

Thus we arrive at the conclusion that the left member of (2) must be less than or equal to M, which contradicts (1).

To be sure, we have only proved that the m.g.f. of X is equal to $\varphi^*(\alpha)$ or $G(\alpha)$ in the open interval $-\alpha_2 < \alpha < \alpha_2$, and not in the corresponding closed interval, as promised. But because of the absolute (and therefore uniform) convergence of the integrals defining $G_n(\alpha)$ and $\varphi(\alpha)$, these functions must be continuous in the closed interval $-\alpha_2 \leqq \alpha \leqq \alpha_2$. Since $\lim_{n\to\infty} G_n(\alpha) = G(\alpha)$ uniformly in this interval, $G(\alpha)$ must also be continuous there. This implies that $\varphi(\alpha)$, the m.g.f. of X, is identically equal to $G(\alpha)$ in the closed interval, and the proof is complete.

It is perhaps worth while to point out explicitly that in the course of the foregoing argument we have proved this proposition:

THEOREM 4. *If a sequence of m.g.f.'s converges in an open interval containing $\alpha = 0$, then it must converge uniformly in every closed subinterval of the open interval, and the limit function is itself a m.g.f.*

REFERENCES

[1] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge, 1937.
[2] D. R. CURTISS, *Analytic Functions of a Complex Variable*, Chicago, 1926.
[3] P. DIENES, *The Taylor Series*, Oxford, 1931.
[4] P. LÉVY, *Théorie de l'Addition des Variables Aléatoires*, Paris, 1937.
[5] D. V. WIDDER, *The Laplace Transform*, Princeton, 1941.

ON THE POWER FUNCTION OF THE ANALYSIS OF VARIANCE TEST[1]

By Abraham Wald

Columbia University

It is known[2] that the general problem of the analysis of variance can be reduced by an orthogonal transformation to the following canonical form: Let the variates $y_1, \cdots, y_p, z_1, \cdots, z_n$ be independently and normally distributed with a common unknown variance σ^2. The mean values of $z_1, \cdots, z_n$ are known to be zero, and the mean values $\eta_1, \cdots, \eta_p$ of the variates $y_1, \cdots, y_p$ are unknown. The canonical form of the analysis of variance test is the test of the hypothesis that

$$\eta_1 = \eta_2 = \cdots = \eta_r = 0 \qquad (r \leq p) \tag{1}$$

where a single observation is made on each of the variates $y_1, \cdots, y_p, z_1, \cdots, z_n$.

In the theory of the analysis of variance the test of the hypothesis (1) is based on the critical region

$$\frac{y_1^2 + \cdots + y_r^2}{z_1^2 + \cdots + z_n^2} \geq c \tag{2}$$

where the constant c is chosen so that the size of the critical region is equal to the level of significance α we wish to have. The critical region (2) is identical with the critical region

$$\frac{y_1^2 + \cdots + y_r^2}{y_1^2 + \cdots + y_r^2 + z_1^2 + \cdots + z_n^2} \geq c' = \frac{c}{c+1}. \tag{3}$$

It is known that the power function of the critical region (3) depends only on the single parameter

$$\lambda = \frac{1}{\sigma^2} \sum_{i=1}^{r} \eta_i^2. \tag{4}$$

Denote the power function of the critical region (3) by $\beta_0(\lambda)$. P. L. Hsu has proved[3] the following optimum property of the region (3): *Let W be a critical region which satisfies the following two conditions*:

(a) *The size of W is equal to the size of the region* (3).

---

[1] Presented at a joint meeting of the Institute of Mathematical Statistics and the American Mathematical Society in New York, December, 1941.

[2] See for instance P. C. Tang, "The power function of the analysis of variance tests," *Stat. Res. Mem.*, Vol. 2, 1938.

[3] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, January, 1941.

(b) *The power function of W depends on the single parameter λ.*
Then $\beta(\lambda) \leq \beta_0(\lambda)$ where $\beta(\lambda)$ denotes the power function of W.

Condition (b) is a serious restriction in Hsu's result. In this paper we shall prove an optimum property of $\beta_0(\lambda)$ where $\beta_0(\lambda)$ is compared with the power function of any other critical region of size equal to that of (3).

For any given values $\eta'_{r+1}, \cdots, \eta'_p, \sigma'$ and λ denote by $S(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)$ the sphere defined by the equations

$$(5) \qquad \eta_1^2 + \cdots + \eta_r^2 = \lambda\sigma'^2; \qquad \eta_i = \eta'_i (i = r+1, \cdots, p); \qquad \sigma = \sigma'.$$

For any region W denote by $\beta_W(\eta_1, \cdots, \eta_p, \sigma)$ the power function of W, i.e. $\beta_W(\eta_1, \cdots, \eta_p, \sigma)$ denotes the probability that the sample point will fall within W calculated under the assumption that $\eta_1, \cdots, \eta_p$ and σ are the true values of the parameters. We will denote by $\gamma_W(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)$ the integral of the power function $\beta_W(\eta'_1, \cdots, \eta'_p, \sigma')$ over the surface $S(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)$ divided by the area of $S(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)$, i.e.

$$(6) \qquad \begin{aligned} &\gamma_W(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda) \\ &= \left[\int_{S(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)} dA\right]^{-1} \int_{S(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda)} \beta_W(\eta'_1, \cdots, \eta'_p, \sigma')\, dA. \end{aligned}$$

We will prove the following

THEOREM: *If W is a critical region of size equal to that of* (3), *i.e.* $\beta_W(0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma) = \beta_0(0)$, *then*

$$(7) \qquad \gamma_W(\eta'_{r+1}, \cdots, \eta'_p, \sigma', \lambda) \leq \beta_0(\lambda)$$

for arbitrary values $\eta'_{r+1}, \cdots, \eta'_p, \sigma'$ and λ.

If W satisfies Hsu's condition (b) then the power function $\beta_W(\eta_1, \cdots, \eta_p, \sigma)$ is constant on the surface $S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)$ and therefore $\gamma_W(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda) = \beta_W(\eta_1, \cdots, \eta_p, \sigma)$. Hence Hsu's result is an immediate consequence of our Theorem.

Denote $|\sqrt{y_1^2 + \cdots + y_r^2 + z_1^2 + \cdots + z_n^2}|$ by t and for any values $a_{r+1}, \cdots, a_p, b$ let $R(a_{r+1}, \cdots, a_p, b)$ be the set of all sample points for which

$$y_i = a_i (i = r+1, \cdots, p) \quad \text{and} \quad t = b.$$

For any region W of the sample space we denote by $W(y_{r+1}, \cdots, y_p, t)$ the common part of W and $R(y_{r+1}, \cdots, y_p, t)$.

In order to prove our Theorem we first show the validity of the following

LEMMA 1: *For any critical region Z there exists a function $\varphi_Z(y_{r+1}, \cdots, y_p, t)$ of the variables $y_{r+1}, \cdots, y_p, t$ such that the critical region Z^* defined by the inequality*

$$y_1^2 + \cdots + y_r^2 \geq \varphi_Z(y_{r+1}, \cdots, y_p, t)$$

satisfies the following two conditions:

(a) $\beta_Z(0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma) = \beta_{Z^*}(0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma)$;

(b) $$\gamma_Z(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda) \leq \gamma_{Z^*}(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda).$$

PROOF. Denote by $P_Z(y_{r+1}, \cdots, y_p, t)$ the conditional probability of $Z(y_{r+1}, \cdots, y_p, t)$ calculated under the condition that the sample point lies in $R(y_{r+1}, \cdots, y_p, t)$ and under the assumption that $\eta_1 = \cdots = \eta_r = 0$ Denote by $F(d, t)$ the conditional probability that

$$y_1^2 + \cdots + y_r^2 \geq d$$

calculated under the condition that the sample point lies in $R(y_{r+1}, \cdots, y_p, t)$ and under the assumption that $\eta_1 = \cdots = \eta_r = 0$. It is easy to verify that the values of $F(d, t)$ and $P_Z(y_{r+1}, \cdots, y_p, t)$ do not depend on the unknown parameters $\eta_{r+1}, \cdots, \eta_p, \sigma$. Since $F(d, t)$ is a continuous function of d and since $F(t^2, t) = 0$, there exists a function $\varphi_Z(y_{r+1}, \cdots, y_p, t)$ such that

$$F[\varphi_Z(y_{r+1}, \cdots, y_p, t), t] = P_Z(y_{r+1}, \cdots, y_p, t).$$

For this function $\varphi_Z(y_{r+1}, \cdots, y_p, t)$ the region Z^* certainly satisfies condition (a) of Lemma 1. We will show that condition (b) is also satisfied. Consider the ratio

$$\frac{\int_{S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}(y_i - \eta_i)^2 - \frac{1}{2\sigma^2}\sum_{\alpha=1}^{n} z_\alpha^2\right] dA}{\exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{r} y_i^2 + \sum_{j=r+1}^{p}(y_j - \eta_j)^2 + \sum_{\alpha=1}^{n} z_\alpha^2\right)\right]} \tag{8}$$

$$= e^{-\frac{1}{2}\lambda} \int_{S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} e^{\sum_{i=1}^{r} y_i \eta_i / \sigma^2} dA.$$

Denote $\left|\sqrt{\sum_{i=1}^{r} y_i^2}\right|$ by r_y. Then we have

$$\int_{S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} e^{\sum_{i=1}^{r} y_i \eta_i / \sigma^2} dA = \int_{(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} e^{\sqrt{\lambda}\, r_y \cos[\alpha(\eta)]/\sigma} dA, \tag{9}$$

where $\alpha(\eta)$ denotes the angle $(0 \leq \alpha(\eta) \leq \pi)$ between the vector y with the components $y_1, \cdots, y_r$ and the vector η with the components $\eta_1, \cdots, \eta_r$. Because of the symmetry of the sphere, the value of the right hand side of (9) is not changed if we substitute $\beta(\eta)$ for $\alpha(\eta)$ where $\beta(\eta)$ denotes the angle $(0 \leq \beta(\eta) \leq \pi)$ between the vector η and an arbitrarily chosen fixed vector u. Hence the value of the right hand side of (9) depends only on r_y, i.e.

$$\int_{S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} e^{\sqrt{\lambda}\, r_y \cos[\alpha(\eta)]/\sigma} dA \tag{10}$$

$$= \int_{S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)} e^{\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma} dA = I(r_y).$$

Now we will show that $I(r_y)$ is a monotonically increasing function of r_y. We have

$$(11)\qquad \frac{dI(r_y)}{dr_y} = \frac{\sqrt{\lambda}}{\sigma}\int_{S(\eta_{r+1},\cdots,\eta_p,\sigma,\lambda)} \cos[\beta(\eta)]\, e^{\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma}\, dA.$$

Denote by ω_1 the subset of $S(\eta_{r+1}, \cdots, \eta_p, \sigma, \lambda)$ in which $0 \le \beta(\eta) \le \frac{\pi}{2}$ and by ω_2 the subset in which $\frac{\pi}{2} \le \beta(\eta) \le \pi$. Because of the symmetry of the sphere we obviously have

$$(12)\qquad \begin{aligned} \int_{\omega_2} \cos[\beta(\eta)] e^{\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma}\, dA &= \int_{\omega_1} \cos[\pi - \beta(\eta)] e^{\sqrt{\lambda}\, r_y \cos[\pi-\beta(\eta)]/\sigma}\, dA \\ &= -\int_{\omega_1} \cos[\beta(\eta)] e^{-\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma}\, dA. \end{aligned}$$

Hence

$$(13)\qquad \frac{dI(r_y)}{dr_y} = \frac{\sqrt{\lambda}}{\sigma}\int_{\omega_1} \cos[\beta(\eta)]\{e^{\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma} - e^{-\sqrt{\lambda}\, r_y \cos[\beta(\eta)]/\sigma}\}\, dA.$$

The right hand side of (13) is positive. Hence $I(r_y)$, and therefore also the left hand side of (8), is a monotonically increasing function of r_y.

Let $P_1(y'_{r+1}, \cdots, y'_p, t', \eta_1, \cdots, \eta_p, \sigma)\, dy_{r+1} \cdots dy_p\, dt$ be the probability that the sample point will fall in the intersection of Z and the set

$$y'_i - \tfrac{1}{2} dy_i \le y_i \le y'_i + \tfrac{1}{2} dy_i (i = r+1, \cdots, p), \quad t' - \tfrac{1}{2} dt \le t \le t' + \tfrac{1}{2} dt$$

Similarly let $P_2(y'_{r+1}, \cdots, y'_p, t', \eta_1, \cdots, \eta_p, \sigma)\, dy_{r+1} \cdots dy_p\, dt$ be the unconditional probability that the sample point will fall in the intersection of Z^* and the set

$$y'_i - \tfrac{1}{2} dy_i \le y_i \le y'_i + \tfrac{1}{2} dy_i (i = r+1, \cdots, p), \quad t' - \tfrac{1}{2} dt \le t \le t' + \tfrac{1}{2} dt.$$

Since the function $\varphi_z(y_{r+1}, \cdots, y_p, t)$ has been defined so that

$$P_z(y_{r+1}, \cdots, y_p, t) = F[\varphi(y_{r+1}, \cdots, y_p, t), t],$$

we obviously have

$$(14)\qquad \begin{aligned} &P_1(y_{r+1}, \cdots, y_p, t, 0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma) \\ &\qquad = P_2(y_{r+1}, \cdots, y_p, t, 0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma). \end{aligned}$$

Using a lemma[4] by Neyman and Pearson, we easily obtain

$$(15)\qquad \begin{aligned} &\int_{S(\eta_{r+1},\cdots,\eta_p,\sigma,\lambda)} P_2(y_{r+1}, \cdots, y_p, t, \eta_1, \cdots, \eta_p, \sigma)\, dA \\ &\qquad \ge \int_{S(\eta_{r+1},\cdots,\eta_p,\sigma,\lambda)} P_1(y_{r+1}, \cdots, y_p, t, \eta_1, \cdots, \eta_p, \sigma)\, dA \end{aligned}$$

[4] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res Mem.*, Vol. 1, London, 1936.

from (14) and the fact that the left hand side of (8) is a monotonically increasing function of $r_y^2 = y_1^2 + \cdots + y_r^2$. Condition (b) is an immediate consequence of (15). Hence Lemma 1 is proved.

For the proof of our theorem we will also need the following

LEMMA 2: *Let $v_1, \cdots, v_k$ be k normally and independently distributed variates with a common variance σ^2. Denote the mean value of v_i by $\alpha_i (i = 1, \cdots, k)$ and let $f(v_1, \cdots, v_k, \sigma)$ be a function of the variables $v_1, \cdots, v_k$ and σ which does not involve the mean values $\alpha_1, \cdots, \alpha_k$. Then, if the expected value of $f(v_1, \cdots, v_k, \sigma)$ is equal to zero, $f(v_1, \cdots, v_k, \sigma)$ is identically equal to zero, except perhaps on a set of measure zero.*

PROOF: Lemma 2 is obviously proved for all values of σ if we prove it for $\sigma = 1$. Hence we will assume that $\sigma = 1$. It is known that a k-variate distribution which has moments equal to those of the joint distribution of $v_1, \cdots, v_k$, must be identical with the joint distribution of $v_1, \cdots, v_k$. That is to say, the joint distribution of $v_1, \cdots, v_k$ is uniquely determined by its moments. Hence if

$$(16) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v_1^{r_1} v_2^{r_2} \cdots v_k^{r_k} g(v_1, \cdots, v_k) e^{-\frac{1}{2}\sum_{i=1}^{k}(v_i - \alpha_i)^2} dv_1 \cdots dv_k = 0$$

for any set $(r_1, \cdots, r_k)$ of non-negative integers, then $g(v_1, \cdots, v_k)$ must be equal to zero except perhaps on a set of measure zero. Now let $f(v_1, \cdots, v_k)$ be a function whose expected value is zero, i.e.

$$(17) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(v_1, \cdots, v_k) e^{-\frac{1}{2}\sum_{i=1}^{k}(v_i - \alpha_i)^2} dv_1 \cdots dv_k = 0$$

identically in $\alpha_1, \cdots, \alpha_k$. From (17) it follows that

$$(18) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(v_1, \cdots, v_k) e^{-\frac{1}{2}\sum_{i=1}^{k} v_i^2 + \sum_{i=1}^{k} \alpha_i v_i} dv_1 \cdots dv_k = 0$$

identically in $\alpha_1, \cdots, \alpha_k$. Differentiating the left hand side of (18) r_1 times with respect to α_1, r_2 times with respect to $\alpha_2, \cdots$, and r_k times with respect to α_k, we obtain

$$(19) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v_1^{r_1} \cdots v_k^{r_k} f(v_1, \cdots, v_k) e^{-\frac{1}{2}\Sigma(v_i - \alpha_i)^2} dv_1 \cdots dv_k = 0.$$

From (16) and (19) it follows that $f(v_1, \cdots, v_k) = 0$. Hence Lemma 2 is proved.

Using Lemmas 1 and 2 we can easily prove our theorem. Because of Lemma 1 we can restrict ourselves to critical regions W which are given by an inequality of the following type

$$y_1^2 + \cdots + y_r^2 \geq \varphi(y_{r+1}, \cdots, y_p, t)$$

where $\varphi(y_{r+1}, \cdots, y_p, t)$ is some function of $y_{r+1}, \cdots, y_p$ and t. The above inequality can be written as

$$\frac{y_1^2 + \cdots + y_r^2}{t^2} \geq \psi(y_{r+1}, \cdots, y_p, t). \tag{20}$$

For any given values of $y_{r+1}, \cdots, y_p, t$ denote by $P(y_{r+1}, \cdots, y_p, t)$ the conditional probability that (20) holds calculated under the assumption that $\eta_1 = \cdots = \eta_r = 0$. It is obvious that $P(y_{r+1}, \cdots, y_p, t)$ does not depend on the unknown parameters $\eta_{r+1}, \cdots, \eta_p, \sigma$. If we denote by W the critical region defined by the inequality (20), we have

$$\begin{aligned} &\beta_W(0, \cdots, 0, \eta_{r+1}, \cdots, \eta_p, \sigma) \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_0^{\infty} P(y_{r+1}, \cdots, y_p, t)\rho_1(y_{r+1}, \cdots, y_p, \eta_{r+1}, \cdots, \eta_p, \sigma) \\ &\qquad \times \rho_2(t, \sigma)\, dy_{r+1} \cdots dy_p\, dt \end{aligned} \tag{21}$$

where $\rho_1(y_{r+1}, \cdots, y_p, \eta_{r+1}, \cdots, \eta_p, \sigma)$ denotes the joint probability density function of $y_{r+1}, \cdots, y_p$ and $\rho_2(t, \sigma)$ denotes the probability density function of t calculated under the assumption that $\eta_1 = \cdots = \eta_r = 0$. In order to satisfy the condition of our Theorem, the function ψ in (20) must be chosen so that

$$\begin{aligned} &\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_0^{\infty} P(y_{r+1}, \cdots, y_p, t)\rho_1(y_{r+1}, \cdots, y_p, \eta_{r+1}, \cdots, \eta_p, \sigma) \\ &\qquad \times \rho_2(t, \sigma)\, dy_{r+1} \cdots dy_p\, dt = \beta_0(0). \end{aligned} \tag{22}$$

Let

$$\int_0^{\infty} P(y_{r+1}, \cdots, y_p, t)\rho_2(t, \sigma)\, dt = Q(y_{r+1}, \cdots, y_p, \sigma). \tag{23}$$

Then we obtain from (22)

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} Q(y_{r+1}, \cdots, y_p, \sigma)\rho_1\, dy_{r+1} \cdots dy_p = \beta_0(0) \tag{24}$$

From (24) and Lemma 2 it follows that

$$Q(y_{r+1}, \cdots, y_p, \sigma) = \beta_0(0) \tag{25}$$

except perhaps on a set of measure zero. From (23), (25) and a result[5] by P. L. Hsu we obtain

$$P(y_{r+1}, \cdots, y_p, t) = \beta_0(0) \tag{26}$$

except perhaps on a set of measure zero.

It follows easily from (26) that $\psi(y_{r+1}, \cdots, y_p, t)$ is equal to a fixed constant except perhaps on a set of measure zero. This proves our Theorem.

[5] P. L. Hsu, "Notes on Hotelling's generalized T," *Annals of Math. Stat*, Vol. 9, p. 237

A NOTE ON THE ESTIMATION OF SOME MEAN VALUES FOR A BIVARIATE DISTRIBUTION

By Edward Paulson[1]

Columbia University

In this paper two problems are discussed which were suggested by the theory of representative sampling [1], but which also occur in several other fields. The first problem is to set up confidence limits for $\frac{m_x}{m_y}$, the ratio of the mean values of the variates x and y. This comes up in the following situation. Let a population π consist of N units $x_1, x_2, \cdots x_N$ and suppose we wish to set up confidence limits for the mean $X = \frac{\sum_{i=1}^{N} x_i}{N}$. Also assume the population π ha[illegible] divided into M groups, let v_j be the number of individuals in the j^{th} grou[illegible]. u_j be the sum of the values of x for the v_j individuals in the j^{th} group, so $X = \frac{u_1 + u_2 \cdots u_M}{v_1 + v_2 \cdots v_M} = \frac{Mm_u}{Mm_v}$. Now if a random sample of n out of the M groups is taken, yielding observations $(u_1, v_1), (u_2, v_2) \cdots (u_n, v_n)$ and N is unknown, the determination of confidence limits for X clearly becomes a special case of the first problem. The distribution of a ratio, discussed by Geary [2], does not seem to be well adapted for this purpose.

The second problem, which is of greater practical interest, arises when we again have a random sample $(u_1, v_1) \cdots (u_n, v_n)$ of n out of M groups and N and M are known. The standard estimate of X that has usually been made is $\hat{X} = \frac{M\bar{u}}{N}$, where $\bar{u} = \frac{\sum_{j=1}^{n} u_j}{n}$. This estimate does not utilize the fact that the n observations on v can be used to increase the precision of the estimate of the numerator of X. This is a special case of problem 2, which we can now formulate as how to best estimate m_x (the mean value of a trait x) both by a point and by an interval, when for each unit in the sample observations both on x and on a correlated variate y are obtainable, and m_y is known a priori. Situations of this type occur fairly often. It is possible to reduce the second problem to the first by using $\frac{\bar{x}}{\bar{y}} \cdot m_y$ as the estimate of m_x, and by multiplying the confidence limits for $\frac{m_x}{m_y}$ by m_y to secure limits for m_x, but this will not usually be the most efficient procedure.

In both problems two cases will be distinguished: (a) when σ_x^2, σ_y^2 and ρ are known a priori, and (b) when they are unknown. To determine confidence

---

[1] Work done under a grant-in-aid from the Carnegie Corporation of New York.

limits for $\frac{m_x}{m_y}$, it will first be assumed that the probability density $f(xy)$ of x and y is

$$f(x, y)$$

$$(1.1) \qquad = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right)+\left(\frac{y-m_y}{\sigma_y}\right)^2\right\}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}.$$

Denote the ratio $\frac{m_x}{m_y}$ by K (assuming $m_y \neq 0$), and suppose it is desired to test the hypothesis that $K = K_0$ on the basis of a sample of n independent observations $(x_1, y_1) \cdots (x_n, y_n)$.

$z_i = x_i - Ky_i$ and $\bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$. Since z is a linear function of x and y it must be normally distributed, and its mean value is obviously zero. Therefore

$$(1.2) \qquad u = \frac{\sqrt{n}\,\bar{z}}{\sigma_z} = \frac{\sqrt{n}\,(\bar{x} - K\bar{y})}{\sqrt{\sigma_x^2 - 2K\rho\sigma_x\sigma_y + K^2\sigma_y^2}}$$

will be normally distributed about zero with unit variance, and the hypothesis is rejected if $|u(K_0)| > u_\alpha$, where $\frac{1}{\sqrt{2\pi}}\int_{u_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\,dt = \frac{1}{2}\alpha$. It is easy to show that this test is equivalent to that based on the likelihood-ratio.

Confidence limits for K would now be given by values of K satisfying the inequality $\left|\frac{\sqrt{n}\,\bar{z}}{\sigma_z}\right| \leq u_\alpha$, provided they always constituted a closed non-empty interval. This is equivalent here to the requirement that K be a real valued monotonic function of u in the interval $-\infty < u < \infty$; this requirement is unfortunately never exactly fulfilled, as can be seen from the graph of (1.2) (in the u, K plane), for the curve has two horizontal asymptotes $u = \pm\frac{\sqrt{n}\,\bar{y}}{\sigma_y}$, and one maximum or minimum point $\left(\text{unless } \frac{\bar{x}}{\bar{y}} = \rho\,\frac{\sigma_x}{\sigma_y}\right)$. However, K will always be a monotonic function of u in the interval $-u_\alpha < u < u_\alpha$ provided $\left|\frac{\sqrt{n}\,\bar{y}}{\sigma_y}\right| > u_\alpha$. Since $m_y \neq 0$, by taking n sufficiently large the probability that $\left|\frac{\sqrt{n}\,\bar{y}}{\sigma_y}\right| < u_\alpha$ can be made arbitrarily small. Moreover, for values of α ordinarily used, in most practical problems the value of $\frac{m_y}{\sigma_y}$ will be such that even for quite small samples the probability $\left|\frac{\sqrt{n}\,\bar{y}}{\sigma_y}\right| < u_\alpha$ (that is, the proba-

bility of getting a sample for which the values of K that are accepted will not form a real interval) will be quite neglgible. For example, let α have the conventional value .05, and suppose $\frac{m_y}{\sigma_y} = 2$; then for $n = 9$, Prob. $\left\{\left|\frac{\sqrt{n}\,\bar{y}}{\sigma_y}\right| < 1.96\right\} < 10^{-4}$ and for $n = 16$, Prob. $\left\{\left|\frac{\sqrt{n}\,\bar{y}}{\sigma_y}\right| < 1.96\right\} < 10^{-9}$. Subject to these rather weak restrictions on the order of magnitude of n and $\frac{m_y}{\sigma_y}$, the confidence limits for K are

$$(1.3) \quad \frac{(n\bar{x}\bar{y} - u_\alpha^2 \rho\sigma_x\sigma_y) \pm \sqrt{(n\bar{x}\bar{y} - u_\alpha^2 \rho\sigma_x\sigma_y)^2 - (n\bar{y}^2 - u_\alpha^2\sigma_y^2)(n\bar{x}^2 - u_\alpha^2\sigma_x^2)}}{n\bar{y}^2 - u_\alpha^2\sigma_y^2}.$$

In case (b) when σ_x^2, σ_y^2, and ρ are unknown, each $z_i = x_i - Ky_i$ is still normally and independently distributed with zero mean and a common variance. It follows that

$$(1.4) \qquad t = \frac{\sqrt{n}\,\bar{z}}{\sqrt{\frac{\Sigma(z - \bar{z})^2}{n - 1}}} = \frac{\sqrt{n}\,(\bar{x} - K\bar{y})}{\sqrt{s_x^2 - 2rs_x s_y K + s_y^2 K^2}}$$

will have Students' distribution with $n - 1$ degrees of freedom. Subject to practically the same restriction as before, the confidence limits for K as determined from (1.4) are

$$(1.5) \quad \frac{(n\bar{x}\bar{y} - t_\alpha^2 rs_x s_y) \pm \sqrt{(n\bar{x}\bar{y} - t_\alpha^2 rs_x s_y)^2 - (n\bar{y}^2 - t_\alpha^2 s_y^2)(n\bar{x}^2 - t_\alpha^2 s_x^2)}}{n\bar{y}^2 - t_\alpha^2 s_y^2}$$

where t_α is the critical value of Students' distribution (for $n - 1$ degrees of freedom) and $s_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$, $s_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n - 1}$, and r is the sample correlation between x and y.

When the distribution of x and y deviates considerably from a bivariate normal one, it would still appear that as a practical matter much the same methods could be used. The basis for this is the fact that there is considerable experimental evidence [3], [4] to show that the distribution of the mean of a sample drawn from any population likely to be encountered in practise will approach normality very rapidly even for n quite small. Hence $\bar{z}$ and u can be regarded as normally distributed for n say >25, and the confidence limits for $\frac{m_x}{m_y}$ will then be given by (1.3); in case (b) a somewhat larger sample is required to diminish the error in estimating σ_z. But for n say >50, t will have a distribution close to normal and the confidence limits for K are given by (1.5) (with t_α replaced by u_α) The statements for the non-normal case appear as a practical matter to also hold when the sample is drawn from a finite population of N

units without replacement if $N - n$ is not too small, provided n is replaced by $n\left(\frac{N-1}{N-n}\right)$, for now $\sigma^2_{(\bar{x}-K\bar{y})} = \frac{1}{n}\left(\frac{N-n}{N-1}\right)[\sigma_x^2 - 2\rho\sigma_x\sigma_y K + \sigma_y^2 K^2]$

In the second problem we again start by assuming the distribution of x and y is given by (1.1). For case (a), m_x is the only unknown parameter If $P = \prod_{i=1}^{n} f(x_i y_i \mid m_x)$ and $\phi = \frac{\partial \log P}{\partial m_x}$, then

$$\phi = \frac{1}{2(1-\rho^2)}\left\{\frac{2\Sigma(x_i - m_x)}{\sigma_x^2} - \frac{2\rho}{\sigma_x\sigma_y}\Sigma(y_i - m_y)\right\},$$

and the maximum likelihood estimate $\hat{m}_1$ of m_x is

$$\hat{m}_1 = \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2}(\bar{y} - m_y), \tag{1.6}$$

where $\sigma_{xy} = \rho\sigma_x\sigma_y$. Also $\hat{m}_1$ is a sufficient statistic, and the confidence interval given by the set of values of m_x satisfying

$$\left|\frac{\sqrt{n}\left[\bar{x} - \frac{\sigma_{xy}}{\sigma_y^2}(\bar{y} - m_y) - m_x\right]}{\sigma_x\sqrt{1-\rho^2}}\right| \leq u_\alpha$$

will be a "shortest unbiased confidence interval" in the sense of Neyman.

Case (b) will be more important, since the exact values of the variances and covariance will usually be unknown. By analogy with (1 6), a similar estimate of m_x for this case is

$$\hat{m}_2 = \bar{x} - \frac{s_{xy}}{s_y^2}(\bar{y} - m_y). \tag{1.7}$$

This is precisely the least square estimate of x, corresponding to $y_i = m_y$, and has been used for this problem before; for example, it is discussed by Cochran [5]. We shall discuss some additional aspects of the problem, and also mention the application to the special case of representative sampling by groups.

When the bivariate distribution of x and y is such that the conditional distribution of each x_i is normal with mean $A + By_i$ and a common variance, then Professor Wald has suggested that exact confidence limits for m_x for small samples can be secured by using the standard methods of the theory of least squares. The resulting confidence limits are easily seen to be

$$\hat{m}_2 \pm \frac{t_\alpha}{\sqrt{n-2}}\sqrt{\lambda},$$

where

$$\lambda = (1 - r^2)\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\frac{1}{n} + \frac{(m_y - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right],$$

and t_α is the critical value of Students' distribution with $n - 2$ degrees of freedom at a level of significance $= \alpha$.

The requirement that the regression of x on y be linear is rather stringent, although it may often be fulfilled, especially in the case of representative sampling mentioned in the opening paragraph. When the regression of x on y is non-linear, the estimate given by (1.7) requires some further justification. Let $U_{ij} = E(x^i y^j)$, where E denotes the mean value, and assume that we have n independent pairs of observations and that the moments U_{10}, U_{01}, U_{11}, U_{20}, U_{02}, U_{40}, U_{04} and U_{22} are all finite. It then follows from a theorem of Doob [6] that $\sqrt{n}(\hat{m}_2 - m_x)$ tends to a limiting distribution with increasing n which is normal with zero mean and variance equal to $\sigma_x^2(1 - \rho^2)$.

The estimate $\bar{x}$ is clearly always less efficient than $\hat{m}_2$ unless $\rho = 0$. The estimate $\frac{\bar{x}}{\bar{y}} \cdot m_y$ is known to have a large sample variance

$$V = \frac{1}{n}\left[\sigma_x^2 - 2\left(\frac{m_x}{m_y}\right)\sigma_{xy} + \left(\frac{m_x}{m_y}\right)^2 \sigma_y^2\right].$$

So $\frac{\bar{x}}{\bar{y}} \cdot m_y$ is always less efficient than $\hat{m}_2$ unless $m_x = \frac{\sigma_{xy}}{\sigma_y^2}\, m_y$, at which point V attains its minimum value $\frac{\sigma_x^2(1 - \rho^2)}{n}$. In fact $\hat{m}_2$ can be easily shown to have an efficiency $\geq$ any other statistic of the class Q, $\left(\text{which includes } \bar{x} \text{ and } \frac{\bar{x}}{\bar{y}}\, m_y\right)$ consisting of all statistics q satisfying two conditions: (1) that $\sqrt{n}(q - m_x)$ have a distribution approaching normality with zero mean and finite variance σ_q^2 and (2) σ_q^2 be independent of the joint density function of x and y, involving only certain of the moments u_{ij}. A rather artificial member of the class Q is $q = \frac{\bar{x} \log \bar{y}}{\log m_y} - \frac{s_x^2}{s_y^2}\left(\sqrt[3]{\bar{y}} - \sqrt[3]{m_y}\right)$. The proof consists merely in observing that if for any bivariate distribution $\sigma_{\hat{m}_2}^2 = \sigma_x^2(1 - \rho^2) > \sigma_q^2$, this would also have to be true when the distribution of x and y is a bivariate normal one, which is impossible, since $\sigma_x^2(1 - \rho^2)$ is then the variance of $\sqrt{n}(\hat{m}_1 - m_x)$, $\hat{m}_1$ being the maximum likelihood statistic.

For moderate values of n, say $n > 100$, fairly exact confidence limits for m_x will be given by $\hat{m}_2 \pm \frac{u_\alpha}{\sqrt{n}} \sqrt{s^2(1 - r^2)}$. When the sample is drawn from a finite population of N units without replacement, the confidence limits for $n > 100$ are $\hat{m}_2 \pm \frac{u_\alpha}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} \sqrt{s_x^2(1 - r^2)}$.

In the problem of estimating $m_x = X$ for the population Π, discussed in the opening paragraph, which consists of N individuals divided into M groups, on the basis of a random sample (u_1, v_1), $(u_2, v_2) \cdots (u_n, v_n)$ of n out of the N

groups, an efficient estimate will be $m' = \dfrac{M\left[\bar{u} - \dfrac{s_{uv}}{s_v^2}\left(\bar{v} - \dfrac{N}{M}\right)\right]}{N}$. The efficiency of m' relative to the conventional estimate $\dfrac{M\bar{u}}{N}$ is $(1 - \rho_{uv}^2)^{-1}$, which ordinarily would seem to be quite large. This is easily extended to the case Π is divided into l strata with M_i groups comprising N_i individuals in the i^{th} stratum, when a random sample of m_i out of the M_i groups in each stratum is taken. Let v_{ij} be the number of individuals in the j^{th} group of the i^{th} stratum and u_{ij} denote the sum of the values of x for these v_{ij} individuals The estimate of m_x becomes

$$m'' = \frac{\sum_{i=1}^{l} M_i \left[\bar{u}_i - \frac{s_{u_i v_i}}{s_{v_i}^2}\left(\bar{v}_i - \frac{N_i}{M_i}\right)\right]}{N}$$

If $\sum_{i=1}^{l} m_i = m$ is fixed, the large sample variance of m'' will be a minimum if m_i is proportional to $M_i\sigma_{u_i}\sqrt{1 - \rho_i^2}$, where ρ_i is the correlation between u and v in the i^{th} stratum.

In conclusion, the writer wishes to thank Professor A. Wald for his advice and encouragement, and Mr. Henry Goldberg for several suggestions.

REFERENCES

[1] J. NEYMAN, "On the two different aspects of the representative method," Journal of the Royal Statistical Society, Vol. 97 (1934), pp. 558–606.

[2] R. C. GEARY, "The frequency distribution of the quotient of two normal variates," *Roy. Stat. Soc. Jour.*, Vol. 93 (1930), pp. 442–446.

[3] W. A. SHEWHART, *Economic Control of Quality of Manufactured Product*, New York, (1931), pp. 182–183.

[4] H. C. CARVER, "Fundamentals in the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 110–112.

[5] W. G. COCHRAN, "The use of the analysis of variance in enumeration by sampling," *Jour. Amer. Stat. Assoc.*, Vol. 34 (1939), pp. 492–510

[6] J. L. DOOB, "The limiting distribution of certain statistics," *Annals of Math Stat*, Vol. 6 (1935), p. 166.

SIGNIFICANCE LEVELS FOR THE RATIO OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE TO THE VARIANCE

By B. I. HART

Ballistic Research Laboratory, Aberdeen Proving Ground

For purposes of practical application in connection with significance tests a tabulation of the argument corresponding to certain percentage points of the probability integral is usually more convenient than that of the probability integral for equal intervals of the argument. A table of probabilities for the

Values of $\frac{\delta^2}{s^2}$ for Different Levels of Significance

	Values of k			Values of k'		
n	$P=.001$	$P=.01$	$P=.05$	$P=.95$	$P=.99$	$P=.999$
4	.7864	.8341	1.0406	4.2927	4.4992	4.5469
5	.5201	.6724	1.0255	3.9745	4.3276	4.4799
6	.4361	.6738	1.0682	3.7318	4.1262	4.3639
7	.4311	.7163	1.0919	3.5748	3.9504	4.2356
8	.4612	.7575	1.1228	3.4486	3.8139	4.1102
9	.4973	.7974	1.1524	3.3476	3.7025	4.0027
10	.5351	.8353	1.1803	3.2642	3.6091	3.9093
11	.5717	.8706	1.2062	3.1938	3.5294	3.8283
12	.6062	.9033	1.2301	3.1335	3.4603	3.7574
13	.6390	.9336	1.2521	3.0812	3.3996	3.6944
14	.6702	.9618	1.2725	3.0352	3.3458	3.6375
15	.6999	.9880	1.2914	2.9943	3.2977	3.5858
16	.7281	1.0124	1.3090	2.9577	3.2543	3.5386
17	.7548	1.0352	1.3253	2.9247	3.2148	3.4952
18	.7801	1.0566	1.3405	2.8948	3.1787	3.4552
19	.8040	1.0766	1.3547	2.8675	3.1456	3.4182
20	.8265	1.0954	1.3680	2.8425	3.1151	3.3840
21	.8477	1.1131	1.3805	2.8195	3.0869	3.3523
22	.8677	1.1298	1.3923	2.7982	3.0607	3.3228
23	.8866	1.1456	1.4035	2.7784	3.0362	3.2953
24	.9045	1.1606	1.4141	2.7599	3.0133	3.2695
25	.9215	1.1748	1.4241	2.7426	2.9919	3.2452
26	.9378	1.1883	1.4336	2.7264	2.9718	3.2222
27	.9535	1.2012	1.4426	2.7112	2.9528	3.2003
28	.9687	1.2135	1.4512	2.6969	2.9348	3.1794
29	.9835	1.2252	1.4594	2.6834	2.9177	3.1594
30	.9978	1.2363	1.4672	2.6707	2.9016	3.1402

	Values of k			Values of k'		
n	$P=.001$	$P=.01$	$P=.05$	$P=.95$	$P=.99$	$P=.999$
31	1.0115	1.2469	1.4746	2.6587	2.8864	3.1210
32	1.0245	1.2570	1.4817	2.6473	2.8720	3.1046
33	1.0369	1.2667	1.4885	2.6365	2.8583	3.0882
34	1.0488	1.2761	1.4951	2.6262	2.8451	3.0725
35	1.0603	1.2852	1.5014	2.6163	2.8324	3.0574
36	1.0714	1.2940	1.5075	2.6068	2.8202	3.0429
37	1.0822	1.3025	1.5135	2.5977	2.8085	3.0289
38	1.0927	1.3108	1.5193	2.5889	2.7973	3.0154
39	1.1029	1.3188	1.5249	2.5804	2.7865	3.0024
40	1.1128	1.3266	1.5304	2.5722	2.7760	2.9898
41	1.1224	1.3342	1.5357	2.5643	2.7658	2.9776
42	1.1317	1.3415	1.5408	2.5567	2.7560	2.9658
43	1.1407	1.3486	1.5458	2.5494	2.7466	2.9545
44	1.1494	1.3554	1.5506	2.5424	2.7376	2.9436
45	1.1577	1.3620	1.5552	2.5357	2.7289	2.9332
46	1.1657	1.3684	1.5596	2.5293	2.7205	2.9232
47	1.1734	1.3745	1.5638	2.5232	2.7125	2.9136
48	1.1807	1.3802	1.5678	2.5173	2.7049	2.9044
49	1.1877	1.3856	1.5716	2.5117	2.6977	2.8956
50	1.1944	1.3907	1.5752	2.5064	2.6908	2.8872
51	1.2010	1.3957	1.5787	2.5013	2.6842	2.8790
52	1.2075	1.4007	1.5822	2.4963	2.6777	2.8709
53	1.2139	1.4057	1.5856	2.4914	2.6712	2.8630
54	1.2202	1.4107	1.5890	2.4866	2.6648	2.8553
55	1.2264	1.4156	1.5923	2.4819	2.6585	2.8477
56	1.2324	1.4203	1.5955	2.4773	2.6524	2.8403
57	1.2383	1.4249	1.5987	2.4728	2.6465	2.8331
58	1.2442	1.4294	1.6019	2.4684	2.6407	2.8260
59	1.2500	1.4339	1.6051	2.4640	2.6350	2.8190
60	1.2558	1.4384	1.6082	2.4596	2.6294	2.8120

ratio of the mean square successive difference δ^2 to the variance s^2, $P\left(\frac{\delta^2}{s^2} < k\right) = \int_0^k \omega(\delta^2/s^2)\, d(\delta^2/s^2)$, where $\omega(\delta^2/s^2)$ is the distribution of δ^2/s^2,[1] has been published recently[2] with k as argument. The following table of values of δ^2/s^2 for $P = .001$, $.01$ and $.05$ has been computed from it by interpolation.

Since the distribution of δ^2/s^2, $\omega(\delta^2/s^2)$, is symmetric[3] about $E(\delta^2/s^2)$, $P(\delta^2/s^2 < k) = P(\delta^2/s^2 > k')$ if $E(\delta^2/s^2) - k = k' - E(\delta^2/s^2)$, where $E(\delta^2/s^2) = 2n/(n-1)$.[3] The upper levels are rarely of practical use, since large values of the ratio, δ^2/s^2, could arise only from a somewhat artificial set of observations, such as alternately high and low values of the observed variable.

The computation of this table of significance levels was made at the suggestion of Lt. Col. L. E. Simon.

[1] For determination of $\omega(\delta^2/s^2)$ cf. John von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Math Stat*, Vol 12 (1941), pp. 367 395.

[2] B I. Hart, "Tabulation of the probabilities for the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13 (1942) p. 213.

[3] Loc. cit.[1] p. 372 for proof of symmetry and evaluation of $E(\delta^2/s^2)$

A CORRECTION

By M. A. Girshick

U. S. Department of Agriculture, Washington

In my article "Notes on the Distribution of Roots of a Polynomial with Random Complex Coefficients" which appeared in the June 1942 issue of the *Annals of Mathematical Statistics*, the symbol $\sum_{p=1}^{n} \sum_{q=p+1}^{n}$ in formulas (13), (14), and (15) should be replaced by $\prod_{p=1}^{n} \prod_{q=p+1}^{n}$.

REPORT OF THE POUGHKEEPSIE MEETING OF THE INSTITUTE

The Fifth Summer Meeting of the Institute of Mathematical Statistics was held at Vassar College, Tuesday and Wednesday, September 8–9, 1942, in conjunction with the meetings of the American Mathematical Society and the Mathematical Association of America. The following fifty-eight members of the Institute attended the meeting:

K. J. Arnold, L. A. Aroian, K. J. Arrow, Walter Bartky, Felix Bernstein, C. I. Bliss, A. H. Bowker, J. H. Bushey, Belle Calderon, B. H. Camp, A. C. Cohen, Jr., A. H Copeland, C. C. Craig, J H. Curtiss, W. E. Deming, J. L. Doob, M. L. Elveback, Willy Feller, M. M. Flood, R. M. Foster, H. A. Freeman, T. N. E. Greville, C. C. Grove, E. J. Gumbel, Edward Helly, G. M. Hopper, Harold Hotelling, Dunham Jackson, R. E. Jolliffe, Irving Kaplansky, Karl Karsten, B F. Kimball, Howard Levene, Eugene Lukacs, H. B. Mann, E. B. Mode, E. C. Molina, F. C. Mosteller, C. R. Mummery, M. L Norden, E. G. Olds, Oystein Ore, Edward Paulson, Selby Robinson, F. E. Satterthwaite, Henry Scheffé, L. E. Simon, Mortimer Spiegelman, Arthur Stein, J. R. Tomlinson, A W. Tucker, J. W. Tukey, D. F. Votaw, Jr , Abraham Wald, S. S. Wilks, E. W. Wilson, Jacob Wolfowitz, L. C. Young.

The opening session, on Tuesday afternoon, was devoted to contributed papers on *Probability and Statistics* and was held jointly with the American Mathematical Society. The Chairman was Professor Cecil C. Craig, University of Michigan, and the following papers were presented:

1. *On the Theory of Testing Composite Hypotheses With One Constraint.*
   Henry Scheffé, Princeton University.
2. *On the Consistency of a Class of Non-parametric Statistics.*
   Jacob Wolfowitz, Staten Island, N. Y.
3. *Graphical Controls Based on Serial Numbers.*
   E. J. Gumbel, New School for Social Research.
4. *Significance Tests for Multivariate Distributions.*
   D. S. Villars, United States Rubber Company. (Introduced by E. G. Olds.)
5. *On the Choice of the Number of Class Intervals in the Application of the Chi-square Test.*
   H. B. Mann and Abraham Wald, Columbia University.
6. *Generalized Poisson Distribution.*
   F. E. Satterthwaite, Aetna Life Insurance Company.
7. *The Relationship of Fisher's z Distribution to Student's τ Distribution.*
   Leo A. Aroian, Hunter College.
8. *On a Statistical Problem Arising in the Classification of an Individual In One of Two Groups.*
   Abraham Wald, Columbia University.
9. *Modern Statistical Methods in Penology.*
   Saly R. R. Struik, Radcliffe College.
   Miriam van Waters, Framingham, Mass.
10. *Regularity of Label-sequences Under Configuration Transformations.*
    T. N. E. Greville, Bureau of the Census.

By Title:

On the Ratio of the Variances of Two Normal Populations.
Henry Scheffé, Princeton University.

Abstracts of these papers follow this report.

On Wednesday morning Professor Harold Hotelling, Columbia University, acted as Chairman of a session on *Stochastic Processes*. The following papers were presented:

1. *Persistence and Recurrence*
   A. H. Copeland, University of Michigan
2. *Definitions and Some Practical Applications*
   Willy Feller, Brown University
3. *General Theory and Applications to Physics.*
   J. L. Doob, University of Illinois

The session on Wednesday afternoon was held jointly with the American Mathematical Society. Lt. Col. Leslie E. Simon, U. S. A., served as Chairman, and the following papers on *The Applicability of Mathematical Statistics to War Efforts* were presented:

1. *Statistical Prediction, With Special Reference to the Problem of Tolerance Limits*
   S. S. Wilks, Princeton University
   Discussant: J. H. Curtiss, Cornell University.
2. *On the Nature of Mathematical Statistics in Quality Control.*
   W. Edwards Deming, Bureau of the Census.
   Discussant: Walter Bartky, University of Chicago.

A meeting of the Board of Directors was held on Tuesday evening. Following the joint dinner on Wednesday evening, a concert was given in Skinner Hall by members of the music department of Vassar College.

EDWIN G. OLDS,
Secretary

ABSTRACTS OF PAPERS

(Presented on September 8, 1942, at the Poughkeepsie meeting of the Institute)

On the Theory of Testing Composite Hypotheses with One Constraint. Henry Scheffé, Princeton University.

A composite hypothesis with one constraint specifies the value of one and only one parameter of a set occurring in a distribution function. The theory of testing such hypothesis is not only of direct interest for many important problems, but is intimately related to Neyman's theory of confidence intervals (*Phil Trans. Roy. Soc. London*, 1937). A method of Neyman (*Bull. Soc. Math France*, 1935) for finding type B regions for testing these hypotheses is extended to the case of any number of nuisance parameters. Type B_1 regions are defined by generalizing the type A_1 regions of Neyman and Pearson (*Stat. Res. Mem.*, 1936) to the case where nuisance parameters are present, and sufficient conditions are found that a type B region be also of type B_1. An interesting moment problem is encountered, in which the admissible functions are not of constant sign, and is solved for the case where the original distribution is multivariate normal

On the Consistency of a Class of Non-Parametric Statistics. J. Wolfowitz, N. Y. City.

Let X and Y be two stochastic variables about whose distribution nothing is known except that they are continuous and let it be required to test whether their distribution functions are the same. Let V be the observed sequence of zeros and ones constructed as described elsewhere (Wald and Wolfowitz, *Annals of Math. Stat.*, Vol. 11 (1940), p. 148). Suppose that the statistic $S(V)$ used to test the hypothesis is of the form $S(V) = \Sigma\varphi(l_j)$, where l_j is the length of the j-th run and $\varphi(x)$ a suitable function defined for all positive integral x. The notion of consistency, originated by Fisher for parametric problems, has already been extended to the non-parametric case (loc. cit., p. 153) The author now proves that, subject to reasonable conditions on $\varphi(x)$ and statistically unimportant restrictions on the alternatives to the null hypothesis, statistics of the type $S(V)$ are consistent. In particular, a statistic discussed by the author (*Annals of Math. Stat.* September, 1942) and for which $\varphi(x) = \log \left(\frac{x^x}{x!}\right)$ belongs to the class covered by the theorem.

Graphical Controls Based on Serial Numbers. E. J. Gumbel, New School for Social Research.

The index m of the observed value x_m ($m = 1, 2, \cdots n$) is called its serial number. A value x of a continuous statistical variable defined by a probability $W(x) = \lambda$ is called a grade (e.g. the median for $\lambda = \frac{1}{2}$). The coordination of serial numbers with grades furnishes two graphical methods for comparing the observations and the theory, namely the equiprobability test based on $m = n\lambda$, and the return periods based on $m = n\lambda + \frac{1}{2}$.

Starting from the distribution of the mth value, we determine the most probable serial number $\tilde{m} = n\lambda + \Delta$, where Δ depends upon the distribution. For a symmetrical distribution, the corrections Δ for two grades defined by λ and $1 - \lambda$, are equal in absolute value and opposite in sign. Then no correction is needed for the median. For an asymmetrical distribution, we calculate the most probable serial number of the mode considered as an mth value. Thus the mode is obtained from the observations through the theory. In this case the mode is not the most precise mth value

If m is of the order $\frac{1}{2}n$, the distribution of the mth value converges towards a normal

distribution with an expectation given by $m = nW(x)$, and a standard deviation $s(x)$, where $s(x)\sqrt{n} = \sqrt{W(x)(1 - W(x))/W(x)}$. By attributing to each theoretical value x its standard deviation, we obtain intervals $x \pm s(x)$ which may be used for the control of the equiprobability test, the comparison of the observed step function with the frequency, and the comparison of the observed with the theoretical return periods. Besides, the standard error of the mth value leads to the precision of the determination of a constant obtained from a grade.

Significance Tests for Multivariate Distributions. D. S. VILLARS, U. S. Rubber Company.

The observed mean of sets of m variates, each normally and independently distributed, is distributed around the population mean according to a χ^2 distribution with m degrees of freedom. The sum of squares of deviations of n observed points from the observed mean is distributed as χ^2 with $m(n - 1)$ degrees of freedom (not with $n - 1$). A much more powerful test for correlation than that by the correlation coefficient is described, which for bivariate distributions, involves comparisons between $n - 1$ and $n - 1$ degrees of freedom. This can be extended to $m - 1$ tests with m variates. Distribution of distance between two means and distribution of fiducial radius is worked out in detail for two variates.

On the Choice of the Number of Class Intervals in the Application of the Chi-Square Test. H. B. MANN and A. WALD, Columbia University.

The distance of two distribution functions is defined as the l.u.b. of the absolute value of the difference between the two cumulative distribution functions. Let $C(\Delta)$ be the class of alternatives with distance $> \Delta$ from the null-hypothesis. Let $f(N, k, \Delta)$ be the g.l.b. of the power with respect to alternatives in $C(\Delta)$ of the chi-square test with sample size N and k equally probable class intervals. A positive integer k is called best with respect to sample size N if there exists a Δ such that $f(N, k, \Delta) = \frac{1}{2}$ and $f(N, k', \Delta) \leqslant \frac{1}{2}$ for every positive integer k'. The authors show that $k_N = \sqrt[5]{\frac{2(N-1)^2}{C^2}}$ where $\frac{1}{\sqrt{2\pi}} \int_c^\infty e^{-\frac{1}{2}x^2}\,dx$ is equal to the size of the critical region, fulfills approximately the conditions of a best k with $\Delta_N = \frac{5}{k_N} - \frac{4}{k_N^2}$ as the corresponding value of Δ. The approximation is shown to be satisfactory for $N \geq 450$ if the 5% level of significance is used and for $N \geq 300$ if the 1% level is used.

Generalized Poisson Distribution. F. E. SATTERTHWAITE, Aetna Life Insurance Company.

In this paper the Poisson distribution is generalized to allow for the assignment of varying weights to a set of events when the number of events follows the Poisson law. The development used brings out the fact that distributions falling in this class do not require that the underlying statistics be homogeneous. The only requirement is that they be independent. Formulas are given for the moments of the generalized distribution as functions of the moments of the underlying distribution of weights. The principles to be observed in the solution of practical problems are outlined.

The Relationship of Fisher's z Distribution to Student's t Distribution. LEO A. AROIAN, Hunter College.

For n_1 and n_2 sufficiently large $W = \frac{1}{\beta}\sqrt{\frac{N}{N+1}}\,z$ is distributed as Student's t with N de-

grees of freedom, $N = n_1 + n_2 - 1$, $\beta^2 = \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$. If the level of significance is α for Student's distribution, the level of significance for z will be $\sqrt{\frac{N}{N+1}}\,\alpha e^{\frac{\beta^2}{3} - \frac{5}{12N}} < \alpha$. As a corollary it follows that the distribution of z approaches normality, $n_1, n_2 \to \infty$, with mean zero and variance $\frac{1}{2}\left(\frac{1}{n_2} + \frac{1}{n_1}\right)$. This simplifies a previous proof of the author. Application of this result is made to finding levels of significance of the z distribution. On the whole R. A. Fisher's formulas for this purpose, n_1 and n_2 large, as modified by W. G. Cochran are superior. The results given by the Fischer-Cochran formulas are compared with those obtained by using the formula recently found by E. Paulson.

On a Statistical Problem Arising in the Classification of an Individual in One of Two Groups. ABRAHAM WALD, Columbia University.

Let π_1 and π_2 be two p-variate normal populations which have a common covariance matrix. A sample of size N_i is drawn from the population $\pi_i (i = 1, 2)$. Denote by $x_{i\alpha}$ the α-th observation on the ith variate in π_1, and by $y_{i\beta}$ the βth observation on the ith variate in π_2. Let $z_i (i = 1, \cdots, p)$ be a single observation on the ith variate drawn from a population π where it is known that π is equal either to π_1 or to π_2. The parameters of the populations π_1 and π_2 are assumed to be unknown. It is shown that for testing the hypothesis $\pi = \pi_1$ a proper critical region is given by $U \geq d$ where $U = \Sigma\Sigma s^{ij} z_i (\bar{y}_j - \bar{x}_j)$, $\| s^{ij} \| = \| s_{ij} \|^{-1}$, $s_{ij} = \left[\sum_\alpha J(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) + \sum_\beta J(y_{i\beta} - \bar{y}_i)(y_{j\beta} - \bar{y}_j)\right] / (N_1 + N_2 - 2)$, $\bar{x}_i = \left(\sum_\alpha J x_{i\alpha}\right) / N_1$, $\bar{y}_i = \left(\sum_\beta J y_{i\beta}\right) / N_2$ and d is a constant. The large sample distribution of U is derived and it is shown that U is a simple function of three angles in the sample space whose exact joint sampling distribution is derived.

Modern Statistical Methods in Penology. SALY R. R. STRUIK, Radcliffe College and MIRIAM VAN WATERS, Massachusetts Reformatory for Women.

In applying statistical methods to penological problems, so far the best known studies have considered 100, 500, or once in England (to refute Lombroso's theory) 1500 cases. But from the correct statistical standpoint, far more cases are needed to establish a law. Over a period of years, an attempt has been made to use statistical methods in the study of penological problems in the Massachusetts Reformatory for Women, but the results will take on real significance and be conclusive only when similar investigations are made all over the United States.

Regularity of Label-Sequences Under Configuration Transformations. T. N. E. GREVILLE, Bureau of the Census.

There is developed a class of transformations on sequences of arbitrary labels in terms of which a wide variety of problems in the theory of probability can be formulated. It is shown that, with mild restrictions on the transformations used and on the measure function assumed on the label-space, almost every label-sequence produces a transform having the frequency distribution expected. The class of transformations considered is shown to include as special cases the four fundamental operations of von Mises: place selection, partition, mixing, and combination.

On the Ratio of the Variances of Two Normal Populations. HENRY SCHEFFÉ, Princeton University.

Let θ be the above ratio. The two problems considered in this paper are the formulation and comparison of (i) significance tests for the hypothesis $\theta = \theta_0$ and (ii) confidence intervals for θ. The paper is divided into two parts: the first is kept on an elementary level and only solutions based on the F distribution are considered. Following various approaches, six tests and corresponding sets of confidence intervals are introduced. It turns out that the limits on the F-distribution which yield an unbiased test are the same as those which yield confidence intervals optimum in a certain intuitive sense. The values of these limits are difficult to compute and some numerical data are given to indicate the loss of efficiency in using instead the easily obtained "equal tails" limits. The second part of the paper is concerned with the existence of common best critical regions and type B_1 regions, and the application of Neyman's theory of confidence intervals. No new tests or confidence intervals not already considered in part I are obtained, but those previously judged best of a very narrow class are now shown to be best of all those based on similar regions of the same size.

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$5.00 per year. Single copies \$1.50. Back numbers are available at \$5.00 per volume, or \$1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879